Sequential Selection under Constraints

Diplomarbeit

vorgelegt von

Mario Stanke
aus
Witzenhausen

angefertigt am

Institut für Mathematische Stochastik
der
Georg-August-Universität zu Göttingen

1999

Contents

1 Introduction

1.1 Abstract and foreword

This thesis deals with a topic in applied probability theory.
We observe a sequence $X_1, X_2, \ldots, X_n$ of nonnegative independent identically distributed random variables sequentially and want to select as many as possible of them so that the sum of the selected random variables does not exceed a given value c. We inspect the random variables in sequence not knowing the X_i's still to come. But we do know the distribution of the random variables. The decision whether we select a random variable must be made at the time it is presented to us. We cannot reject a random variable later once we have chosen to select it, neither can we select one that we have rejected earlier. We call a rule saying which random variables to select an *online selection policy* if it meets above requirement.
The problem will be to examine the largest possible expected number of selected random variables within all online selection policies and find a good policy in that sense.
Various papers have been written on that topic, most of them referring to [3] – a paper by Coffman, Flatto and Weber – who first examined the problem for a distribution function F of the X_i of the form $F(x) \sim Ax^\alpha$ as $x \to 0$. The authors also note that, if the online restriction in above problem was dropped, and one could simply inspect all X_i's at the beginning and then select the smallest ones first, one could not do much better than with the online restriction. The advantage such a 'prophet' would have becomes asymptotically negligible as n becomes large.
Later, their results have been generalized by Rhee and Talagrand ([14]) to arbitrary distribution functions F. Gnedin ([8]) has introduced and examined a generalization of this problem. In this variant the number of random variables n itself is a random variable N with known distribution and which is independent of the X_i's. Again, we sequentially inspect $X_1, X_2, \ldots, X_N$ but don't know how many random variables there will be. In this problem, too, the advantage of knowing the sequence $X_1, X_2, \ldots$ (but not N) in advance does not lead to an asymptotically better result than the best online policy gives.
The thesis in hand examines both the problem with fixed n and with random N for d-dimensional random vectors $X_1, X_2, \ldots$ in the positive orthant of $\mathbb{R}^d$. The sum restriction is replaced by the restriction that the sum of the selected random variables must be less than or equal to a constant $\mathbf{c} \in \mathbb{R}^d$ in each coordinate.
Here is an – admittedly not very realistic – example. A butcher has d different kinds of meat. On a certain day N customers come to her shop, the j-th customer demanding $X_j^{(i)}$ grams of meat sort i. She knows from her experience the distribution of N and of the customer demands X_j. But on that day it is foreseeable that she won't be able to serve all the customer demands. She only has c_i grams of meat sort i. Being afraid of losing customers , she wants to annoy as few of them as possible. She decides to try to serve as many customer demands *completely* as possible. Who should she serve?
Another weakness of this example is that she might use her knowledge of the remaining time until her shop closes to predict the number of customers still to come which is not allowed in our model. But I think the example is still realistic enough to illustrate the

features of a heuristic policy. She will serve customers who demand little of a kind what she has little left of and she will rather serve the same demand in the beginning than later. Chapter 2 deals with the problem for fixed n. First the structure of an optimal online selection policy is examined and its existence is proven. But the exact solution to this problem for a given nontrivial instance, i.e. distribution and n, seems intractable. We instead examine the asymptotics of the maximal expected number of selected variables when $n \to \infty$ for an arbitrary continuous distribution of the X_i's. As upper bound on this quantity will serve the best possible performance of the prophet as mentioned above. In order to establish the lower bound – by actually giving a simple online policy that asymptotically achieves optimality – we first examine another problem. Given a measure on the positive orthant we need to maximize the volume V of a region R in the positive orthant under the restriction that $V\cdot$ (barycenter of R) stays in each coordinate below a given value.
In chapter 3 we examine the maximal expected number of selected random variables for random N asymptotically. What asymptotically here exactly means ('N large') will have to be specified. This problem is more general than the one with fixed n but we can only solve it for a smaller class of distributions.
In section 4.1 we give an upper bound on the expected number of random variables a prophet can select in the one-dimensional case. This will also serve as an upper bound on the best possible performance of online selection policies. We will use this result in chapters 2 and 3.

I would like to thank Dr. habil. Alexander V. Gnedin, who always took the time to help me when I had problems.

1.2 Notations and definitions

Most of this thesis deals with the d-dimensional real vector space $\mathbb{R}^d$.
For $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, $\mathbf{x} = (x_1, x_2, \ldots, x_d)$, $\quad \mathbf{y} = (y_1, y_2, \ldots, y_d)$ we will write

$$\mathbf{x} \leq \mathbf{y} :\Longleftrightarrow x_i \leq y_i \quad \text{for } i = 1, 2, \ldots, d$$

and

$$\mathbf{x} < \mathbf{y} :\Longleftrightarrow x_i < y_i \quad \text{for } i = 1, 2, \ldots, d.$$

If $\mathbf{z} \in \mathbb{R}^d$ we will assume that $\mathbf{z} = (z_1, z_2, \ldots, z_d)$ without saying so. The same holds if $\mathbf{z}$ is a function with values in $\mathbb{R}^d$. If a symbol already has a subscript – e.g. we have a sequence $(\mathbf{z}_n)_n$ with $\mathbf{z}_n \in \mathbb{R}^d$ – we will write superscripts to denote the coordinates: $\mathbf{z}_n = (z_n^{(1)}, z_n^{(2)}, \ldots, z_n^{(d)})$.
$\|\mathbf{x}\|_p = (\sum_{i=1}^d x_i^p)^{1/p}$ denotes the p-Norm of $\mathbf{x}$ $(1 \leq p < \infty)$ and $\|\mathbf{x}\|_\infty := \max_i |x_i|$.
$\langle \mathbf{x}, \mathbf{y} \rangle$ denotes the Euclidean scalar (inner) product of $\mathbf{x}$ and $\mathbf{y}$.
Write $\mathbf{1} := (1, 1, \ldots, 1) \in \mathbb{R}^d$ and $\mathbf{0} := (0, 0, \ldots, 0) \in \mathbb{R}^d$. Let $\mathbb{R}_+^d := \{\mathbf{x} \in \mathbb{R}^d \,|\, \mathbf{x} \geq \mathbf{0}\}$.
For $A, B \subset \mathbb{R}^d$ write $A \leq B$ if $\mathbf{a} \leq \mathbf{b}$ for all $\mathbf{a} \in A$ and $\mathbf{b} \in B$. Correspondingly, for the other comparing operators. We will also write $\mathbf{x} \leq A$ meaning $\mathbf{x} \leq \mathbf{a}$ for all $\mathbf{a} \in A$.

Definition 1.1 A set $A \subset \mathbb{R}_+^d$ is a *lower-layer* if $\mathbf{x} \in A$, $\mathbf{0} \leq \mathbf{y} \leq \mathbf{x}$ implies $\mathbf{y} \in A$.

For two subsets A, B of the same space M let $A \diamond B := A \setminus B \cup B \setminus A$ denote the symmetric difference. Let $A^c = M \setminus A$ be the complement of A in M. If M is a topological space let A° denote the interior of A.
For two real functions $f, g : M \to \mathbb{R}$, and $y \in M$ we will write

$$f(x) \sim g(x) \quad (\text{as } x \to y) \quad :\Longleftrightarrow \lim_{x \to y} \frac{f(x)}{g(x)} = 1$$

Usually M will be $\mathbb{R}$ or $\mathbb{N}$. Then $y = \infty$ is allowed and always meant when nothing else is stated. That is, we will then just write $f \sim g$. We will also write

$$f \gtrsim g :\Longleftrightarrow \liminf_{x \to \infty} \frac{f(x)}{g(x)} \geq 1 \quad \text{and} \quad f \lesssim g :\Longleftrightarrow \limsup_{x \to \infty} \frac{f(x)}{g(x)} \leq 1.$$

Note that $\gtrsim$ and $\lesssim$ are transitive and that $f \gtrsim g, f \lesssim g$ implies $f \sim g$.
Further, for $x, y \in \mathbb{R}$, $x^+ := \max\{x, 0\}$ is the positive part of x and $x \wedge y$ denotes the minimum of x and y.
Let $\mathcal{B}(\mathbb{R}^d)$ be the Borel σ-algebra on $\mathbb{R}^d$ (generated by the open sets). And for measurable $M \subset \mathbb{R}^d$, i.e. $M \in \mathcal{B}(\mathbb{R}^d)$ we will write $\mathcal{B}(M)$ for the restriction of $\mathcal{B}(\mathbb{R}^d)$ to M. When we take $A \subset M$ we implicitly assume that A is measurable. λ will always denote the Lebesgue measure – usually on $\mathbb{R}^d$. For a given measure μ on M let $\mathcal{N}$ denote the set of sets of measure 0. If f is a measurable mapping from M to $\mathbb{R}$ we will write $\|f\|_p := (\int_M |f|^p \, d\mu)^{1/p}$ for the L_p-norm. Also $\|f\|_\infty = \text{ess sup } f$.
Let $Poi(s)$ denote the Poisson distribution with parameter s.

1.3 Formulation of the general problem

Let $\mathbf{0} < \mathbf{c} \in \mathbb{R}^d$, and let $Q = [\mathbf{0}, \mathbf{c}]^d$ be the d-dimensional interval with endpoints $\mathbf{0}$ and $\mathbf{c}$. Let $X_1, X_2, \ldots \geq \mathbf{0}$ be independent identically distributed d-dimensional random vectors with law μ on $\mathbb{R}^d_+$. Let $\mathbf{X} = (X_1, X_2, \ldots)$. We will speak of the X_i's as sizes of *items* and of subintervals of Q as *space*. Let N be a positive integer-valued random variable that is independent of X. N is the number of items that are at disposal to be packed. Let ν be the distribution of N.
All random variables are assumed to be defined on some common probability space $(\Omega, \mathcal{A}, P)$.

Definition 1.2 A *selection policy* is a function
$\Psi = (\Psi_1, \Psi_2, \ldots) : \prod_{\mathbb{N}} \mathbb{R}^d_+ \to \{0,1\}^{\mathbb{N}}$.

Definition 1.3 An *online selection policy* is a selection policy $\Psi = (\Psi_1, \Psi_2, \ldots)$ where $\mathbf{x} = (x_1, x_2, \ldots) \in \prod_{\mathbb{N}} \mathbb{R}^d_+$ is mapped to $\Psi(\mathbf{x}) = (\Psi_1(x), \Psi_2(x), \ldots)$ and

$$\Psi_j(\mathbf{x}) = \Psi_j(x_1, \ldots, x_j) \qquad (j = 1, 2, \ldots),$$

that is, Ψ_j is a function of $x_1, x_2, \ldots, x_j$ only.

Throughout this thesis a selection policy Ψ will be regarded as a function of the random sequence $\mathbf{X}$ and we will usually write Ψ instead of $\Psi(\mathbf{X})$ and Ψ_j for $\Psi_j(X)$. We say that item j of size X_j is selected by Ψ if $\Psi_j(X_1, \ldots, X_j) = 1$ and it is rejected if $\Psi_j(X_1, \ldots, X_j) = 0$. Note that for an online selection policy, Ψ_j depends only on $X_1, X_2, \ldots, X_j$ and therefore the random variables Ψ_j, X_i and N are independent for $i > j$.

We consider the restriction that the sum of the selected variables must stay within Q. We call those policies *admissible* that satisfy the sum constraint

$$\sum_{j=1}^{\infty} \Psi_j(\mathbf{x}) x_j \leq \mathbf{c} \qquad \text{for all } \mathbf{x} \in \prod_{\mathbb{N}} \mathbb{R}^d_+. \tag{1}$$

We are interested in the expected number of selected variables

$$\mathcal{E}(\Psi) := E \sum_{j=1}^{N} \Psi_j \tag{2}$$

and want to maximize it.

Let $\mathcal{P}$ be the set of all admissible *online* selection policies.
And let $\mathcal{S}$ be the set of all admissible selection policies.

Define

$$\text{Opt} := \sup_{\Psi \in \mathcal{P}} \mathcal{E}(\Psi)$$

and

$$\text{Proph} := \sup_{\Psi \in \mathcal{S}} \mathcal{E}(\Psi),$$

the maximal expected number of selected items within the respective class of policies.

Interpretation: *N is interpreted as the number of items. In (2) only the items selected up to time N are counted, so we might as well interpret it as if there were no more items. In $\mathcal{P}$ the decision whether to select X_j or not depends only on X_j and the 'past': $X_1, X_2, \ldots, X_{j-1}$. The items come one after the other and we have to decide* online, *i.e. without knowing the 'future' and without revoking a decision we have made before. Our decision can also depend on everything that is not random: The distribution of $\mathbf{X}$ and N. But it cannot depend on $N(\omega)$ or $X_{j+1}(\omega), X_{j+2}(\omega), \ldots$.*
In $\mathcal{S}$ the decision can depend on the whole seqence $X_1, X_2, \ldots$. We imagine a prophet who is given the task of selecting the items. The prophet knows the sizes of all the items in advance. His decision also can depend on the distribution of N. But it cannot depend on $N(\omega)$ itself. So, the prophet is clairvoyant only to the sizes of the items, not to their number N. We have chosen this partially clairvoyance because it will turn out that it does not help the prophet much when the number of items becomes large. If the prophet also knew N in advance, he could do significantly better even for large numbers of items.

This paper mainly deals with *Opt* and we have introduced *Proph* only to establish an upper bound for *Opt*: As $\mathcal{P} \subset \mathcal{S}$, clearly

$$\text{Opt} \leq \text{Proph}.$$

As *Opt* and *Proph* depend on the distribution of $\mathbf{X}$ and N and on $\mathbf{c}$ we will write $\text{Opt} = \text{Opt}(\mu, \mathbf{c}, \nu)$ and $\text{Proph} = \text{Proph}(\mu, \mathbf{c}, \nu)$.
We will sometimes use this definition for a finite measure μ which has total mass $p := \mu(\mathbb{R}_+^d) < 1$ and therefore is not a probability measure. This should be interpreted in the following way. With probability $1 - p$ the item is discarded right away and cannot be selected. This is the same as defining the probability measure $\mu' := \mu + (1-p)\varepsilon_{\infty}$, where ε_{∞} denotes the Dirac measure with unit mass in some large enough point ∞ that does not 'fit' in Q anyway, and then defining $\text{Opt}(\mu, \mathbf{c}, \nu) := \text{Opt}(\mu', \mathbf{c}, \nu)$ and $\text{Proph}(\mu, \mathbf{c}, \nu) := \text{Proph}(\mu', \mathbf{c}, \nu)$.

In chapter 2 we will examine $\text{Opt}(\mu, \mathbf{c}, \nu)$ when $N \equiv n$ is not random. And in chapter 3 we will treat a more general case for ν but with stronger assumptions on μ.

2 Sequential Selection out of n Random Vectors under a Sum Constraint

In this chapter we consider the case when the total number of available items N is constant:

$$N \equiv n$$

or in other words ν is the Dirac measure with unit mass in n.
We have to use an online selection policy Ψ to select items from $X_1, X_2, \ldots, X_n$ under the sum restriction (1). As the distribution of N (given by ν) is determined by n, we will now write

$$\mathrm{Opt}_n(\mu, \mathbf{c}) := \mathrm{Opt}(\mu, \mathbf{c}, \nu)$$

for the maximal expected number of selected items.

In section 2.1 we will give a recursion formula for $\mathrm{Opt}_n(\mu, \mathbf{c})$ and derive that Opt_n is monotone in $\mathbf{c}$ and μ in some sense.
In section 2.2 we will give some preliminary results on $\mathrm{Opt}_n(\mu, \mathbf{c})$ for $n \to \infty$ before introducing an asymptotically optimal selection policy in section 2.4.

2.1 The exact solution

We can give a recursion formula (or Bellman equation) for the maximal expected number of selected items $\mathrm{Opt}_n(\mu, \mathbf{c})$. Unfortunately, the computation of its solution typically seems to be intractable. Nevertheless, we can derive some theoretical results about $\mathrm{Opt}_n(\mu, \mathbf{c})$ from the recursion. Also, the existence of an optimal strategy will follow by induction on n.
Fix the probability measure μ. If $n = 1$, an optimal policy is to select X_1 if we can, i.e. if $X_1 \leq \mathbf{c}$. Then

$$\mathrm{Opt}_1(\mu, \mathbf{c}) = \mu([\mathbf{0}, \mathbf{c}]). \tag{3}$$

Now, let $\mathrm{Opt}_n(\mu, \mathbf{c})$ already be defined for all $\mathbf{c}$ and let there be a policy which attains this optimum.
Suppose that the first out of $n+1$ random variables $X_1, X_2, \ldots, X_{n+1}$ has been observed and that $X_1 = \mathbf{x}$. If $\mathbf{x} \not\leq \mathbf{c}$ we can't select it. In this case – or if we choose to reject it – we are left with the problem of packing n items into the space $\mathbf{c}$. For this problem we have an optimal policy which selects $\mathrm{Opt}_n(\mu, \mathbf{c})$ items on average. But if $\mathbf{x} \in [\mathbf{0}, \mathbf{c}]$ and we choose to select it we are left with space $\mathbf{c} - \mathbf{x}$ and n more items. Here, too, we have an optimal policy with $\mathrm{Opt}_n(\mu, \mathbf{c} - \mathbf{x})$ as expected number of selected items (in addition to the first one).
For $\mathbf{x} \in [\mathbf{0}, \mathbf{c}]$ our decision will be based on which one of the values $\mathrm{Opt}_n(\mu, \mathbf{c})$ and $1 + \mathrm{Opt}_n(\mu, \mathbf{c} - \mathbf{x})$ is larger. An optimal strategy is to select X_1 iff

$$1 + \mathrm{Opt}_n(\mu, \mathbf{c} - \mathbf{x}) \geq \mathrm{Opt}_n(\mu, \mathbf{c}) \tag{4}$$

and then apply the corresponding optimal policy to the subproblem with n items. And if we set $\mathrm{Opt}_n(\mu, \mathbf{c}) := -\infty$ for $\mathbf{c} \not\geq \mathbf{0}$ the expected number of selected items is

$$\max\{\mathrm{Opt}_n(\mu, \mathbf{c}), 1 + \mathrm{Opt}_n(\mu, \mathbf{c} - \mathbf{x})\}$$

for all $\mathbf{x} \in \mathbb{R}_+^d$. Averaging in $\mathbf{x}$, we get the recursion

$$\mathrm{Opt}_{n+1}(\mu, \mathbf{c}) = \int\limits_{\mathbb{R}_+^d} \max\{\mathrm{Opt}_n(\mu, \mathbf{c}), 1 + \mathrm{Opt}_n(\mu, \mathbf{c} - \mathbf{x})\}\, d\mu(\mathbf{x}) \tag{5}$$

Now, we would like to compare $\mathrm{Opt}_n(\mu, \mathbf{c})$ for different μ's. Intuitively, if we have two measures μ and μ' and independent identically distributed random variables $X_1, X_2, \ldots$ distributed according to μ and $X_1', X_2', \ldots$ distributed according to μ' such that the X_i are in some sense smaller than the X_i' then we should be able to pack more of the X_i than of the X_i'.

Definition 2.1 For two finite measures μ and μ' on $\mathbb{R}_+^d$ introduce the *stochastic order*

$$\mu \succeq \mu' \quad :\Longleftrightarrow \quad \int h\, d\mu \geq \int h\, d\mu' \quad \text{for all decreasing and } \mu\text{-and } \mu'\text{-integrable functions } h : \mathbb{R}_+^d \to \mathbb{R}_+.$$

Here, *decreasing* means $h(\mathbf{x}) \leq h(\mathbf{y})$ whenever $\mathbf{x} \geq \mathbf{y}$.

Lemma 2.2

$$\mu \succeq \mu' \quad \Longleftrightarrow \quad \mu(A) \geq \mu'(A) \text{ for every lower-layer } A.$$

Proof.
"$\Rightarrow$" Take $h = \mathbb{1}_A$, which is decreasing because A is a lower layer.
So $\mu(A) = \int h\, d\mu \geq \int h\, d\mu' = \mu(A')$.
"$\Leftarrow$" Let h be a decreasing and μ-and μ'-integrable function from $\mathbb{R}_+^d$ to $\mathbb{R}_+$ and let $\mu(A) \geq \mu'(A)$ for every lower-layer A. Transforming measure (compare [1]: Bauer, Maßtheorie, Satz 23.8) we get

$$\int\limits_{\mathbb{R}_+^d} h(\mathbf{x})\, d\mu(\mathbf{x}) = \int\limits_0^\infty \mu(\{h \geq t\})\, dt.$$

As h is decreasing, the sets $\{h \geq t\}$ are lower-layers and $\mu(h \geq t) \geq \mu'(h \geq t)$. We get

$$\int h\, d\mu = \int_0^\infty \mu(\{h \geq t\})\, dt \geq \int_0^\infty \mu'(\{h \geq t\})\, dt = \int h\, d\mu'.$$

□

Remark: *In the one-dimensional case $d = 1$, $\mu \succeq \mu'$ is equivalent to the distribution function of μ being pointwise greater than or equal the distribution function of μ', since the one-dimensional lower-layers are intervals $[0, a)$ or $[0, a]$. For $d \geq 2$ this is a stronger assumption than this inequality between the distribution functions.*

Lemma 2.3 (monotonicity lemma)

(1) $\mathrm{Opt}_n(\mu, \mathbf{c})$ is monotone in $\mathbf{c}$, i.e.

$$\mathbf{c} \geq \mathbf{c}' \Rightarrow \mathrm{Opt}_n(\mu, \mathbf{c}) \geq \mathrm{Opt}_n(\mu, \mathbf{c}') \tag{6}$$

for all probability measures μ on Q.

(2) $\mathrm{Opt}_n(\mu, \mathbf{c})$ is monotone in μ, i.e.

$$\mu \succeq \mu' \Rightarrow \mathrm{Opt}_n(\mu, \mathbf{c}) \geq \mathrm{Opt}_n(\mu', \mathbf{c}) \tag{7}$$

for all $\mathbf{c} \in \mathbb{R}^d_+$.

Remark: *Of course,* $\mathrm{Opt}_n(\mu, \mathbf{c})$ *is monotone in n, too.*
The lemma remains valid for a measure μ with $\mu(R^d_+) < 1$.

Proof.

(1) Straightforward induction on n using equation (5)
(The case $\mathbf{c} - \mathbf{x} \not\geq \mathbf{0}$ has to be treated separately.)

(2) Base case: $n = 1$, follows from (3).
Induction hypothesis: (7) holds for all $\mathbf{c} \in \mathbb{R}^d_+$.

$$\mathrm{Opt}_{n+1}(\mu, \mathbf{c}) = \int\limits_{\mathbb{R}^d_+} h(\mathbf{x}) d\mu(\mathbf{x}), \tag{8}$$

where

$$h(\mathbf{x}) := \max\{\mathrm{Opt}_n(\mu, \mathbf{c}), 1 + \mathrm{Opt}_n(\mu, \mathbf{c} - \mathbf{x})\}$$

is a function from Q to the nonnegative real numbers. g is a decreasing because of (6). By the definition of the stochastic order we have $\int h \, d\mu \geq \int h \, d\mu'$. We conclude

$$\begin{aligned}
\mathrm{Opt}_{n+1}(\mu, \mathbf{c}) &\geq \int_{\mathbb{R}^d_+} \max\{\mathrm{Opt}_n(\mu, \mathbf{c}), 1 + \mathrm{Opt}_n(\mu, \mathbf{c} - \mathbf{x})\} \, d\mu'(\mathbf{x}) \\
&\geq \int_{\mathbb{R}^d_+} \max\{\mathrm{Opt}_n(\mu', \mathbf{c}), 1 + \mathrm{Opt}_n(\mu', \mathbf{c} - \mathbf{x})\} \, d\mu'(\mathbf{x}) \qquad (9) \\
&= \mathrm{Opt}_{n+1}(\mu', \mathbf{c}),
\end{aligned}$$

where the step (9) follows from the induction hypothesis.

□

Now, we see that the optimal policy described above is of 'threshold' type. Let $A_n(\mathbf{c})$ be the set of all $\mathbf{x}$ such that (4) holds. Call $A_n(\mathbf{c})$ the *acceptance region*. Because of (6) the acceptance region $A_n(\mathbf{c})$ is a lower-layer. We will state this as

Theorem 2.4 (Existence of an optimal strategy) There is an optimal policy Ψ with $\mathcal{E}(\Psi) = \mathrm{Opt}_n(\mu, \mathbf{c})$, such that the acceptance regions $\{x_j \in Q \mid \Psi_j(x_1, \ldots, x_j) = 1\}$ are lower-layers which depend on $x_1, x_2, \ldots, x_{j-1}$ only through the sum of the variables selected so far: $\sum\limits_{k=1}^{j-1} \Psi_k x_k$.

2.2 Preliminary asymptotical results

In general, we cannot determine $\mathrm{Opt}_n(\mu, \mathbf{c})$ exactly. Instead, we we will now focus on its asymptotic behavior when $n \to \infty$.
Let us first make a standardization of the units. Instead of using a general $\mathbf{c} = (c_1, \ldots, c_d)$ as upper bound on the sum of the selected items, we will use $\mathbf{1} = (1, 1, \ldots, 1)$. We can always transform the problem with general $\mathbf{c}$ to the one with $\mathbf{c} = \mathbf{1}$ by considering items with sizes X'_j measured on a different scale:

$$X' = (X'_1, X'_2, \ldots), \qquad X_j^{(i)\prime} := X_j^{(i)}/c_i, \qquad i = 1, 2, \ldots, d, \quad j = 1, 2, \ldots.$$

From now on let Q denote the unit cube $[\mathbf{0}, \mathbf{1}]^d$. We will now write $\mathrm{Opt}_n(\mu)$ for $\mathrm{Opt}_n(\mu, \mathbf{c})$.

2.2.1 A coarse look on the asymptotics

Lemma 2.5 Let $\mu(A) > 0$ for any neighborhood A of $\mathbf{0}$. Then

$$\lim_{n\to\infty} \mathrm{Opt}_n(\mu) = \infty.$$

Proof. $\mathrm{Opt}_n(\mu)$ is monotonically increasing in n so the limit exists. Let $M > 1$ be arbitrary, let $A := [\mathbf{0}, \frac{1}{M}\mathbf{1}]$ and consider the policy Ψ which selects all items with sizes in A unless the sum of the selected items would exceed $\mathbf{1}$. Then

$$\mathrm{Opt}_n(\mu) \geq \mathcal{E}(\Psi) \geq M\, P\Big(\sum_{j=1}^{n} \mathbb{1}_{\{X_j \in A\}} \geq M\Big) \to M,$$

since $\sum_{j=1}^{n} \mathbb{1}_{\{X_j \in A\}}$ is binomially distributed with parameters n and $\mu(A) > 0$.
Therefore $\lim\limits_{n\to\infty} \mathrm{Opt}_n(\mu) \geq M$ and the claim follows as M was arbitrary.

□

If there is a neighborhood A of $\mathbf{0}$ in Q such that $\mu(A) = 0$, we can almost surely only select a bounded number of items. As this case doesn't seem very interesting in our asymptotical analysis we will exclude it from further consideration. From now on let

$$\mu(A) > 0 \text{ for any neighborhood } A \text{ of } \mathbf{0}. \tag{10}$$

For the asymptotic behavior of $\mathrm{Opt}_n(\mu)$ only the values of $\mu(A)$ for neighborhoods of $\mathbf{0}$ play a role.

Lemma 2.6 If μ_1 and μ_2 are equal in a neighborhood U of $\mathbf{0}$, i.e. $\mu_1(A) = \mu_2(A)$ for all $A \subset U$, then

$$\mathrm{Opt}_n(\mu_1) \sim \mathrm{Opt}_n(\mu_2).$$

Proof. Let $\mathcal{E}_\mu(\Psi)$ denote the expected number of selected items of policy Ψ with respect to a measure μ.
Let $\Psi \in \mathcal{P}$ be an optimal policy with respect to the measure μ_1. The number of items selected by Ψ which have a size that is not in U is bounded by some constant M, which depends only on U.
Define Ψ' by $\Psi'_j := \Psi_j \mathbb{1}_{\{X_j \in U\}}$. Ψ' selects all the items $X_j \in U$ that would be selected by Ψ. Therefore $\mathcal{E}_{\mu_1}(\Psi') \geq \mathcal{E}_{\mu_1}(\Psi) - M$.
Although Ψ' does not select items not in U, it still can depend on those items because Ψ depended on them. But there is a policy Φ which is at least as good as Ψ' with respect to μ_1 and does neither select items not in U *nor* depend on them.
We can see that by letting $\mathcal{P}'$ be the set of all admissible online policies which only select items j which have a size $X_j \in U$. Like in the preceeding section one can see that for any measure there is an optimal policy Φ under all policies in $\mathcal{P}'$, too. And Φ depends on $X_1, X_2, \cdots, X_j$ only through the sum of the items selected up to time j. In particular Φ_j does not depend on the sizes of the items not in U which haven't been selected.
As $\Psi' \in \mathcal{P}$ we get $\mathcal{E}_{\mu_1}(\Psi') \leq \mathcal{E}_{\mu_1}(\Phi)$. But Φ does not depend on the sizes of the items not in U and since μ_1 and μ_2 are equal on U we have $\mathcal{E}_{\mu_1}(\Phi) = \mathcal{E}_{\mu_2}(\Phi)$.
We conclude

$$\mathrm{Opt}_n(\mu_1) = \mathcal{E}_{\mu_1}(\Psi) \leq \mathcal{E}_{\mu_1}(\Psi') + M \leq \mathcal{E}_{\mu_1}(\Phi) + M = \mathcal{E}_{\mu_2}(\Phi) + M \leq \mathrm{Opt}_n(\mu_2) + M$$

as Φ is only suboptimal for μ_2.
Since the situation is symmetric in μ_1 and μ_2 we get

$$|\mathrm{Opt}_n(\mu_1) - \mathrm{Opt}_n(\mu_2)| \leq M.$$

Dividing this inequality by $\mathrm{Opt}_n(\mu_2)$ and letting $n \to \infty$ now gives the result since $\mathrm{Opt}_n(\mu_2) \to \infty$ by lemma 2.5.

□

2.2.2 Poissonization of the selection problem

So far we had a fixed number $n \in \mathbb{N}$ of items at our disposal to be packed. It will be of some use later to make this discrete parameter n to a continuous parameter by the means of Poissonization.

Consider a homogeneous Poisson process $\{L(t), t \geq 0\}$ on $\mathbb{R}_+$ with rate 1 that is independent of the sequence of items $\mathbf{X}$. Let T_j be the arrival time of the j-th event in the Poisson process. We observe $(X_1, T_1), (X_2, T_2), \ldots$ sequentially and select the X_j's online under the constraint that the sum of the selected variables must not exceed $\mathbf{c}$. But we cannot select X_j if $T_j > t$. In other words the selection process stops at time t. Our decision whether we select X_j (i.e. $\Psi_j = 1$) now can depend on $(X_1, T_1), (X_2, T_2), \ldots, (X_j, T_j)$.
An admissible online selection policy here is a function
$\Psi = (\Psi_1, \Psi_2, \ldots) : \quad \prod_{\mathbb{N}} (\mathbb{R}_+^d \times \mathbb{R}_+) \to \{0,1\}^{\mathbb{N}}$,
such that for $\mathbf{z} = ((x_1, t_1), (x_2, t_2), \ldots)$ in the domain of Ψ

$$\Psi_j(\mathbf{z}) = \Psi_j((x_1, t_1), (x_2, t_2), \ldots, (x_j, t_j)) \qquad (j = 1, 2, \ldots)$$

and the sum constraint

$$\sum_{j=1}^{\infty} \Psi_j(\mathbf{z})\, x_j \leq \mathbf{c}$$

is satisfied for all $\mathbf{z}$ in the domain of Ψ. Define

$$\mathcal{E}(\Psi) := \mathrm{E} \sum_{j:\, T_j \leq t} \Psi_j.$$

And let

$$u(\mu, \mathbf{c}, t) := \sup \mathcal{E}(\Psi) \tag{11}$$

where Ψ ranges in the supremum over all admissible online selection policies. If $\mathbf{c} = \mathbf{1}$ we will just write $u(\mu, t)$ and usually $u(t)$ when it is clear which measure we use.

At the time the j-th event of the Poisson process occurs the j-th item X_j is presented to us an we have to decide whether to select it. We neither know the sizes X_i for $i > j$ nor do we know how many events will occur in the future. But we do know the current time and the time left up to t. We know the past and the distributions.
$u(t)$ is the maximal expected number of selected items up to time t.

$N := L(t)$, the number of events up to time t, is Poisson distributed with mean t. As the variance of N is also t, we get by Chebyshev's inequality that $P(|N/t - 1| > \varepsilon) \leq 1/(\varepsilon^2 t) \to 0$ as $t \to \infty$. So for large t, N is distributed sharpely about its mean t and we can hope that the maximal expected number of selected items then is similar in the two cases: $N \equiv n$ and the Poissonized problem with time horizont n.

What is so useful about $u(t)$ instead of Opt_n is that the first is accessible by analytical methods. We will make use of that in section 2.5.1.

Lemma 2.7

$$\frac{u(t)}{t} \text{ is decreasing in } t. \tag{12}$$

Proof. Let $0 \leq p \leq 1$ and suppose each of the items j $(j = 1, 2, \ldots)$ is marked 'valid' with probability p independently of the rest of the process. And when counting the number of selected items only *valid* items are considered.
We will make the following Gedanken experiment. There are two players A and B, who have to select the items up to time t coming from a Poisson process with rate 1. The players play their policies independently from each other. Player A just plays an optimal policy which maximizes the expected total number of selected items, valid or not. (She doesn't even know that there are invalid items.)
Player B sees the mark on an item before she must decide if she selects it.
Clearly, the maximal expected number of selected *valid* items of player B will be as least as large as the expected number of selected *valid* items of player A, because B could choose to apply the same strategy as A and ignore the marks.
The expected number of selected items of player A is $u(t)$ and thus her expected number of selected *valid* items is $p\,u(t)$, because the marks were given independently of the rest of the process with probability p.
For Player B, there is no sense to pick invalid items because they only occupy space and do not add to the performance. She will simply ignore any invalid items and consider only the valid items in her selection process. The thinned counting process of the valid items is again a Poisson process on $[0, t]$. It has rate p. This selection problem is equivalent to the one for a Poisson process with rate 1 on $[0, pt]$ - we just rescale the time. So the expected number of selected valid items of player A is $u(pt)$.
We conclude

$$u(pt) \geq p\,u(t),$$

which implies

$$\frac{u(pt)}{pt} \geq \frac{u(t)}{t}.$$

As we have that for all $t > 0$ and $0 \leq p \leq 1$ the claim follows.

□

We are now ready to prove the conjecture from the introduction to this subsection. Fix μ and let

$$a_n := \mathrm{Opt}_n(\mu).$$

Theorem 2.8 The maximal expected number of selected items is asymptotically the same in the cases $N \equiv n$ and the Poissonized problem with time horizont n:

$$a_n \sim u(n) \tag{13}$$

Proof. We prove (13) in two parts:
(a) $a_n \lesssim u(n)$ and
(b) $a_n \gtrsim u(n)$.
For part (a) take $\varepsilon > 0$ and set $t := n(1+\varepsilon)$. Consider an optimal strategy Ψ for the problem with fixed number n of items. Let $N \sim Poi(t)$, then we get with Chebyshev's inequality

$$P(N < n) \leq P(|N - t| \geq n\varepsilon) \leq n(1-\varepsilon)/(n\varepsilon)^2 \to 0 \qquad \text{as } n \to \infty. \tag{14}$$

In the Poissonized packing problem with $N \sim Poi(t)$ one could apply this policy Ψ to the first n incoming items and ignore all items thereafter. If $N < n$ the selection simply stops with the last item. This admissible strategy will get at least $a_n P(N \geq n)$ items on the average but is still suboptimal in the Poissonized problem. So $a_n P(N \geq n) \leq u(t)$. We get

$$\begin{aligned} \frac{a_n}{u(n)} &= \frac{a_n}{u(t)} \frac{u(t)}{u(n)} \\ &\leq \frac{1}{P(N \geq n)} \frac{t}{n} & , \text{since } \frac{u(t)}{u(n)} \leq \frac{t}{n} \text{ by lemma 2.7} \\ &= \frac{1+\varepsilon}{P(N \geq n)} \\ &\to 1+\varepsilon & \text{by (14)} \end{aligned}$$

So $\limsup_n \frac{a_n}{u(n)} \leq 1 + \varepsilon$ for all $\varepsilon > 0$, and (a) follows.

For (b) we need to bound $u(t)$ appropriately from above. This time, let $t := n(1-\varepsilon)$ and $0 < \varepsilon < 1$. Let N be the number of items in the Poissonized problem with time t. Suppose now, we knew N in advance. N is still random but the value is revealed to us in advance. Since the item sizes are independent of N, the maximal expected number of selected items would be $\mathrm{E}\, a_N$. Because we would have more information than in the Poissonized packing problem, we would be able to pack at least as much on the average. So

$$u(t) \leq \mathrm{E}\, a_N.$$

Now, conditioning on whether $N \leq n$ we get

$$\begin{aligned} u(t) &\leq \mathrm{E}\,[a_N \,|\, N \leq n]\; P(N \leq n) + \mathrm{E}\,[a_N \,|\, N > n]\; P(N > n) \\ &\leq a_n\; P(N \leq n) + \mathrm{E}\,[N \,|\, N > n]\; P(N > n) \qquad , \text{since } a_k \uparrow \text{ and } a_k \leq k. \end{aligned} \tag{15}$$

We have

$$P(N > n) = \mathrm{O}(1/n) \tag{16}$$

just as in part (a).
To bound $\mathrm{E}[N \mid N > n]$ (we only need here that it is $\mathrm{O}(n)$) note that, as N is a Poisson random variable with mean t, we have $P(N = k) = \frac{t}{k} P(N = k-1)$ for $k = 1, 2, \ldots$. So the same holds for the conditional distribution of N given $N > n$:

$$P(N = k \mid N > n) = \frac{t}{k} P(N = k-1 \mid N > n) \qquad k > n+1.$$

Set $q := 1 - \varepsilon$ and observe that $t/k \leq 1 - \varepsilon = q$ for $k > n+1$ to get

$$P(N = k \mid N > n) \leq q P(N = k-1 \mid N > n) \qquad k > n+1.$$

Iterating this inequality we get

$$P(N = n+1+k \mid N > n) \leq q^k P(N = n+1 \mid N > n) \qquad k = 0, 1, 2, \ldots. \tag{17}$$

We can now compare the conditional probability distribution of $N - (n+1)$ given $N > n$ with a geometric distribution with parameter $p = 1 - q = \varepsilon$

$$\begin{aligned}
\mathrm{E}[N - (n+1) \mid N > n] &= \sum_{k=1}^{\infty} k P(N - (n+1) = k \mid N > n) \\
&\leq \sum_{k=1}^{\infty} k q^k P(N = n+1 \mid N > n) \qquad \text{(by (17))} \\
&\leq \sum_{k=1}^{\infty} k q^k = \frac{1}{p^2} = \frac{1}{\varepsilon^2}
\end{aligned}$$

So $\mathrm{E}[N \mid N > n] \leq n + 1 + 1/\varepsilon^2$ indeed is $\mathrm{O}(n)$. Together with (16) and (15) this yields

$$\frac{a_n}{u(t)} \geq \big(1 - \frac{\mathrm{O}(1)}{u(t)}\big) \Big/ P(N \leq n).$$

$u(t) \to \infty$ using part (a) and $a_k \to \infty$ (holds by lemma 2.5). So

$$\liminf_n \frac{a_n}{u(t)} \geq 1$$

and

$$\liminf_n \frac{a_n}{u(n)} = \frac{a_n}{u(t)} \frac{u(t)}{u(n)} \geq \liminf_n \frac{u(t)}{u(n)} \overset{\text{lemma 2.7}}{\geq} \lim_{n \to \infty} \frac{t}{n} = 1 - \varepsilon.$$

As $\varepsilon > 0$ was arbitrarily small, (b) follows.

□

With lemma 2.6 we have seen that the asymptotic behavior of $\mathrm{Opt}_n(\mu)$ depends only on μ in a neighborhood of $\mathbf{0}$. We will now see that the asymptotical behavior of $\mathrm{Opt}_n(\mu)$ even only depends on 'the asymptotic behavior of μ' in a neighborhood of $\mathbf{0}$.

Theorem 2.9 Let $\mu_1 = f_1\lambda$ and $\mu_2 = f_2\lambda$ be two probability measures on $\mathbb{R}_+^d$ with densities f_1 and f_2 with respect to the Lebesgue measure λ. And let

$$f_1(\mathbf{x}) \sim f_2(\mathbf{x}) \qquad \text{as } \|\mathbf{x}\| \to 0.$$

Then

$$\mathrm{Opt}_n(\mu_1) \sim \mathrm{Opt}_n(\mu_2) \qquad (n \to \infty).$$

Proof. Let $0 < a < 1$ be given. (We will let $a \uparrow 1$ later). Since $f_1(\mathbf{x}) \sim f_2(\mathbf{x})$ there is a neighborhood U of $\mathbf{0}$ such that $f_2(\mathbf{x}) \geq af_1(\mathbf{x})$ for all $\mathbf{x} \in U$. Define the measure

$$\mu := (f_1 \mathbb{1}_U)\lambda,$$

i.e. μ is like μ_1 but 'restricted' to U. Then $\mu_2 = f_2\lambda \succ (af_1\mathbb{1}_U)\lambda = a\mu$. So we can apply the second part of the monotonicity lemma 2.3 to get

$$\begin{aligned} \mathrm{Opt}_n(\mu_2) &\geq \mathrm{Opt}_n(a\mu) \\ &\sim u(a\mu, n) \qquad \text{because of theorem 2.8.} \end{aligned}$$

Now, observe that in the Poissonized problem multiplying the measure by a factor $0 < a < 1$ has the same effect on the maximal expected number of selected items as scaling the time with that factor:

$$u(a\mu, \mathbf{c}, t) = (\mu, \mathbf{c}, at). \tag{18}$$

Using theorem 4.6 about 'thinning' a Poisson process we see that in both cases the items i such that $X_i \in Q$ come from a homogenuous Poisson process and their number is Poisson distributed with mean $at\mu(Q)$. The conditional distribution of X_i given $X_i \in Q$ is the same, too, in both cases, namely $\frac{1}{\mu(Q)}\mu$. And these two distributions solely determine the maximal expected number of selected items. Apply this (for $\mathbf{c} = \mathbf{1}$) to get

$$\begin{aligned} \mathrm{Opt}_n(\mu_2) &\gtrsim u(\mu, an) \\ &\geq au(\mu, n) && \text{(by lemma 2.7)} \\ &\sim a\mathrm{Opt}_n(\mu) && \text{(by theorem 2.8)} \\ &\sim a\mathrm{Opt}_n(\mu_1) && \text{(by lemma 2.6)} \end{aligned}$$

So we have $\mathrm{Opt}_n(\mu_2) \gtrsim a\,\mathrm{Opt}_n(\mu_1)$ for a arbitrarily close to 1, which implies $\mathrm{Opt}_n(\mu_2) \gtrsim \mathrm{Opt}_n(\mu_1)$. By symmetry we conclude $\mathrm{Opt}_n(\mu_1) \gtrsim \mathrm{Opt}_n(\mu_2)$ and the claim follows.

□

2.3 Sets of maximal volume under certain restrictions

Our aim still is to determine $\mathrm{Opt}_n(\mu)$ asymptotically. And give a simple policy which asymptotically achieves the optimum. We have seen in section 2.1 that for the optimal strategy there are acceptance regions $A_k(\mathbf{c})$, where k is the number of items still to inspect and $\mathbf{c}$ is the space left. The simplest strategy seems to choose a fixed acceptance region A and accept all the items with sizes in A as long as allowed by the sum constraint. We call this a stationary strategy. If there was no constraint the expected number of selected items would be $n\mu(A)$ and the expected space needed would be $n\mathrm{E}\left[X_1 \mathbb{1}_A(X_1)\right]$. It seems natural to try to use an A such that $n\mu(A)$ is maximal under all A's such that the expected space needed is less than or equal to $\mathbf{1}$. It will turn out that such an A (which depends on n) indeed gives an asymptotically optimal admissible stragety when $n \to \infty$.
In this section we will deal with the problem of determining the shape of A. Although only very little of the results of this section is actually needed for the rest of the chapter (the optimality of the region is not needed) we included it. Partly, because the subject deserves its own interest and partly because the proof of the result we do need later follows the same line.

2.3.1 Problem and notations

We are given a probability measure μ on the d-dimensional unit cube $Q = [0,1]^d$.

Definition 2.10 By a *simplicial section* we mean a set $\{\mathbf{x} \in Q \mid \langle \mathbf{x}, \boldsymbol{\theta} \rangle \leq 1\}$, where $\boldsymbol{\theta} \geq \mathbf{0}$.

Figure 5 on page 31 shows one.
Assume the random vector Z has law μ. For a measurable set $A \subset Q$ let

$$\mathbf{g}(A) := E[Z \, \mathbb{1}_A(Z)] = \int_A \mathbf{x} \, d\mu(\mathbf{x}).$$

And let $\mathbf{c}(A)$ for $\mu(A) > 0$ denote the barycenter (center of gravity) of A with respect to μ, i.e.

$$\mathbf{c}(A) := E[Z \mid Z \in A] = \frac{\mathbf{g}(A)}{\mu(A)} = \frac{1}{\mu(A)} \int_A \mathbf{x} \, d\mu(\mathbf{x}).$$

The problem of this section will be an optimization problem:

$$(P_1) \qquad \text{maximize } \mu(A) \quad \text{on } \{A \subset Q \mid \mathbf{g}(A) \leq \boldsymbol{\rho}, \quad A \text{ measurable}\}$$

for some $\boldsymbol{\rho} > \mathbf{0}$.
The solution of this problem will also give us a solution to the problem

$$(P_2) \qquad \text{maximize } \mu(A) \quad \text{on } \{A \subset Q \mid \mathbf{c}(A) \leq \boldsymbol{\tau}, \quad A \text{ measurable}\}$$

for some $\boldsymbol{\tau} > \mathbf{0}$.
Problem (P_2) has also been treated in [12]:
Mallows, Nair, Shepp and Vardi (1985), *Optimal Sequential Selection of Secretaries*
Unfortunately, their proof, attributed to Andrew Odlyzko, seems to have a gap when $d > 2$. We will use a different approach here.

2.3.2 Counterexamples

In the paper mentioned above are the following two lemmas

Lemma 2.11 For any A there is a lower-layer A' with $\mathbf{c}(A') \leq \mathbf{c}(A)$ and $\mu(A') \geq \mu(A)$.

Lemma 2.12 For any lower-layer A there is a simplicial section A' with $\mathbf{c}(A') \leq \mathbf{c}(A)$ and $\mu(A') \geq \mu(A)$.

The cited paper claims that they are right for an arbitrary distribution μ. However, both statements fail in the case where the measures have atoms. The same statements with **c** replaced by **g** are false in general, too, for arbitrary distributions μ. It will turn out later that they are indeed correct for a continuous measure μ (with **c** and with **g**).
Figures 1 and 2 show counterexamples for a non-continuous measure.

Let A contain **a** and **c** but not **b**. Let μ be the probability measure which puts the masses $\varepsilon, 1-2\varepsilon, \varepsilon$ onto the points $\mathbf{a}, \mathbf{b}$ and **c**, respectively, for some small ε. Then there is no lower-layer A' with $\mu(A') \geq \mu(A)$ and $\mathbf{c}(A') \leq \mathbf{c}(A)$.
Every lower-layer A' with $\mu(A') \geq \mu(A)$ would have to contain **b** and thus had a larger center of gravity than A: $\mathbf{c}(A') > \mathbf{c}(A)$, which is a contradiction and also shows that then
$\mathbf{g}(A') = \mu(A')\mathbf{c}(A') > \mu(A)\mathbf{c}(A) = \mathbf{g}(A)$.
So there is no lower-layer A' with $\mu(A') \geq \mu(A)$ and $\mathbf{g}(A') \leq \mathbf{g}(A)$, either.

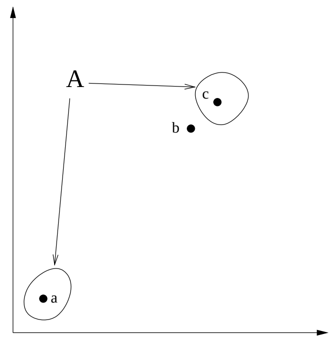


Figure 1: counterexample 1

Let A be the square shown in the figure. Let μ be the probability measure which puts the masses $\frac{1}{3}(1-\varepsilon)$ onto the points $\mathbf{a}, \mathbf{b}$ and **c**, and the mass ε on point **d**, for some small ε.
Then there is no simplicial section A' with $\mu(A') \geq \mu(A)$ and $\mathbf{c}(A') \leq \mathbf{c}(A)$.
Every simplicial section A' with $\mu(A') \geq \mu(A)$ would have to contain **a** and at least one of **b** and **c** and thus had a center of gravity $\mathbf{c}(A')$ which is in at least one coordinate larger than $\mathbf{c}(A)$ (which is very close to **a** when ε is small): $\mathbf{c}(A') \not\leq \mathbf{c}(A)$.
This also shows that then
$\mathbf{g}(A') = \mu(A')\,\mathbf{c}(A') \not\leq \mu(A)\,\mathbf{c}(A) = \mathbf{g}(A)$.
So there is no simplicial section A with $\mu(A') \geq \mu(A)$ and $\mathbf{g}(A') \leq \mathbf{g}(A)$, either.

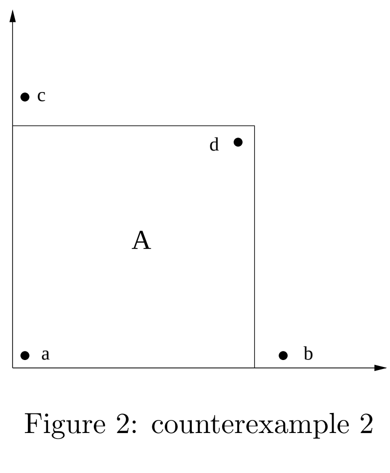


Figure 2: counterexample 2

2.3.3 Existence

In this subsection we will prove that above optimization problems have a solution. Let μ be a continuous measure with density f. The idea is to take a sequence of subsets A_n of Q which satisfy the restriction and approach the optimal value, then to approximate these with lower-layers L_n. And then use the 'smoothness' of lower-layers to prove there is a 'convergent' subsequence of this sequence (L_n). The limes will then be a solution.

Introduce the pseudometric d_μ on the set of Borel-measurable sets

$$d_\mu(A,B) := \mu(A \diamond B) \qquad\qquad A,B \in \mathcal{B}(\mathbb{R}^d),$$

where $\diamond$ denotes the symmetric difference.

Lemma 2.13 The mappings $\mathbf{g} : \mathcal{B}(\mathbb{R}^d_+) \to \mathbb{R}^d$ and $\mathbf{c} : \mathcal{B}(\mathbb{R}^d_+) \setminus \mathcal{N} \to \mathbb{R}^d$ are continuous with respect to d_μ.

Proof. Let $A, B \in \mathcal{B}(\mathbb{R}^d_+)$ and let $i \in \{1, 2, \cdots, d\}$. Then

$$|g_i(A) - g_i(B)| = \Big| \int\limits_{A\setminus B} x_i \, d\mu(\mathbf{x}) - \int\limits_{B\setminus A} x_i \, d\mu(\mathbf{x}) \Big| \le \int\limits_{A\diamond B} |x_i| \, d\mu(\mathbf{x}) \le \mu(A \diamond B) = d_\mu(A,B).$$

So

$$\|\mathbf{g}(A) - \mathbf{g}(B)\|_\infty \le d_\mu(A,B) \tag{19}$$

and thus $\mathbf{g}$ is continuous. The continuity of $\mathbf{c}(A) = \frac{1}{\mu(A)}\mathbf{g}(A)$ follows since the mapping μ is of course continuous with respect to d_μ, too.

□

Lemma 2.14 Let μ be an absolutely continuous measure on Q.

(a) For every $A \subset Q$ and every $\varepsilon > 0$ there is a lower-layer $L \subset Q$ such that $\mu(L) \ge \mu(A) - \varepsilon$ and $\mathbf{g}(L) \le \mathbf{g}(A) + \varepsilon \mathbf{1}$.

(b) For every $A \subset Q$ with $\mu(A) \ne 0$ and every $\varepsilon > 0$ there is a lower-layer $L \subset Q$ such that $\mu(L) \ge \mu(A) - \varepsilon$ and $\mathbf{c}(L) \le \mathbf{c}(A) + \varepsilon \mathbf{1}$.

Proof. We will first show that (a) implies (b). And then prove (a).
(a) $\Rightarrow$ (b): Let (a) be true, let $\varepsilon > 0$ be given and L be so that $\mu(L) \ge \mu(A) - \delta$ and $\mathbf{g}(L) \le \mathbf{g}(A) + \delta\mathbf{1}$. We will choose δ with $\varepsilon \ge \delta > 0$ later. Then

$$\mathbf{c}(L) = \frac{\mathbf{g}(L)}{\mu(L)} \le \frac{\mathbf{g}(A) + \delta\mathbf{1}}{\mu(A) - \delta} \overset{*}{\le} \frac{\mathbf{g}(A)}{\mu(A)} + \varepsilon\mathbf{1} = \mathbf{c}(A) + \varepsilon\mathbf{1},$$

where inequality $(*)$ holds when we choose $\delta > 0$ small enough, which is possible since $\mu(A) > 0$ and the left hand side of $(*)$ tends to $\mathbf{g}(A)/\mu(A)$ as $\delta \to 0$. Also $\mu(L) \ge \mu(A) - \varepsilon$ as $\delta \le \varepsilon$.

(a) We will first construct a B with $\mu(B) = \mu(A)$ and $\mathbf{g}(B) \le \mathbf{g}(A)$ such that B is close to a lower-layer with respect to d_μ. The idea is to turn A into a set B by making 'successive improvements'.
Let $A_1 := A$ and let A_k already be defined for some $k \ge 1$. For $\mathbf{y} \in Q$ let

$$\delta_{\mathbf{y}}(A_k) := \min\{\mu(A_k^c \cap [\mathbf{0}, \mathbf{y}]),\ \mu(A_k \cap [\mathbf{y}, \mathbf{1}])\}. \tag{20}$$

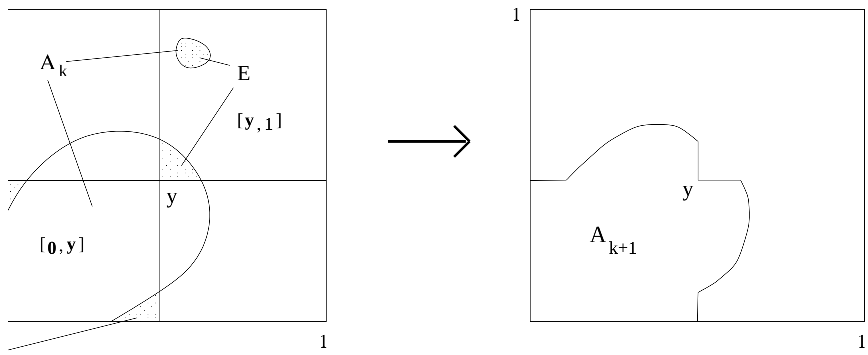


Figure 3: Example for the 'improvement' of A_k

See also figure 3. And choose $\mathbf{y}$ so that $\delta_{\mathbf{y}}(A_k)$ is maximal. This is possible since $\delta_{\mathbf{y}}$ is continuous in $\mathbf{y}$ and Q is compact. Now, we choose $D \subset A_k^c \cap [\mathbf{0}, \mathbf{y}])$ and $E \subset A_k \cap [\mathbf{y}, \mathbf{1}]$ such that $\mu(D) = \mu(E) = \delta_{\mathbf{y}}(A_k)$. Again, this is possible because μ is continuous.
Finally, define $A_{k+1} := (A_k \cup D) \setminus E$.
Graphically, this means we have moved the mass $\delta_{\mathbf{y}}(A_k)$ down (in each coordinate) from E to D. Since D and A_k are disjoint and $E \subset A_k$ we get $\mu(A_{k+1}) = \mu(A_k)$. We also get

$$\mathbf{g}(A_k) - \mathbf{g}(A_{k+1}) = \int_E \mathbf{x}\, d\mathbf{x} - \int_D \mathbf{x}\, d\mathbf{x} \ge \mathbf{0},$$

since $D \le \mathbf{y} \le E$ and $\mu(E) = \mu(D)$.
We would like to continue this process up to some n such that $\delta_{\mathbf{y}}(A_n) < \varepsilon'$ for some given, arbitrarily small $\varepsilon' > 0$. It is true that $\delta_{\mathbf{y}}(A_{k+1}) \le \delta_{\mathbf{y}}(A_k)$ but this alone doesn't mean that $\delta_{\mathbf{y}}(A_n) < \varepsilon'$, eventually. Using the continuity of μ we will show that if $\delta_{\mathbf{y}}(A_k) \ge \varepsilon'$ then $\langle \mathbf{g}, \mathbf{1}\rangle = g_1 + \cdots + g_d$ will decrease each step by an amount that is bounded from below by a constant. But as $\langle \mathbf{g}, \mathbf{1}\rangle$ is bounded from below by 0 this means that $\delta_{\mathbf{y}}(A_n) < \varepsilon'$ for some n.
Let $\varepsilon' > 0$ be given and let $\delta_{\mathbf{y}}(A_k) \ge \varepsilon'$. As μ is continuous with respect to Lebesgue measure λ there is a constant $\delta > 0$ such that $\mu(C) < \varepsilon'/2$ for all $C \subset Q$ with $\lambda(C) < \delta$. Let Δ_s be the simplex $\{\mathbf{x} \ge \mathbf{y} \,|\, \langle \mathbf{x}, \mathbf{1}\rangle \le \langle \mathbf{y}, \mathbf{1}\rangle + s\}$. Now, choose $s > 0$ so small that $\lambda(\Delta_s) < \delta$ and observe that it doesn't depend on $\mathbf{y}$. Then $\mu(\Delta_s) < \varepsilon'/2$.
We conclude that

$$\mu(E \cap \Delta_s^c) = \mu(E) - \mu(E \cap \Delta_s) \ge \delta_y - \mu(\Delta_s) \ge \varepsilon'/2. \tag{21}$$

Now, we can bound

$$\begin{aligned}
\begin{array}{l}\langle \mathbf{g}(A_k), \mathbf{1}\rangle - \\ \langle \mathbf{g}(A_{k+1}), \mathbf{1}\rangle\end{array} &= \int_E \langle \mathbf{x}, \mathbf{1}\rangle \; d\mu(\mathbf{x}) - \int_D \langle \mathbf{x}, \mathbf{1}\rangle \; d\mu(\mathbf{x}) \\
&\geq \int\limits_{E\cap\Delta_s^c} \langle \mathbf{x}, \mathbf{1}\rangle \; d\mu(\mathbf{x}) + \int\limits_{E\cap\Delta_s} \langle \mathbf{x}, \mathbf{1}\rangle \; d\mu(\mathbf{x}) - \mu(D) \langle \mathbf{y}, \mathbf{1}\rangle \qquad (22) \\
&\geq \mu(E\cap\Delta_s^c)(\langle \mathbf{y}, \mathbf{1}\rangle + s) + \mu(E\cap\Delta_s) \langle \mathbf{y}, \mathbf{1}\rangle - \mu(D) \langle \mathbf{y}, \mathbf{1}\rangle \qquad (23) \\
&\geq \underbrace{(\mu(E) - \mu(D))}_{=0} \langle \mathbf{y}, \mathbf{1}\rangle + s\mu(E\cap\Delta_s^c) \\
&\geq s\varepsilon'/2,
\end{aligned}$$

where (22) holds since $D \leq \mathbf{y}$ and (23) holds since for $\mathbf{x} \in \{E\cap\Delta_s^c\}$ we have $\langle \mathbf{x}, \mathbf{1}\rangle > \langle \mathbf{y}, \mathbf{1}\rangle + s$ by definition of Δ_s and since $E \geq \mathbf{y}$.
The constant $s\varepsilon'/2 > 0$ doesn't depend on A_k. So we can't have $\delta_{\mathbf{y}}(A_k) \geq \varepsilon'$ for all k, because then we would have $\langle \mathbf{g}(A_k), \mathbf{1}\rangle < 0$ eventually, which is impossible. We must have $\delta_y(A_n) < \varepsilon'$ for n large enough.
Set $B := A_n$. Then $\mu(B) = \mu(A)$, $\mathbf{g}(B) \leq \mathbf{g}(A)$ and $\delta_{\mathbf{y}}(B) < \varepsilon'$ for all $\mathbf{y} \in Q$.

Now, let $\varepsilon > 0$ be given. Let $m \in \mathbb{N}$ be so large that $\lambda(C) \leq d/m$ implies $\mu(C) < \varepsilon/3$, which again is possible since μ is continuous. Let $\varepsilon' := \frac{\varepsilon}{3m^d}$ and B be as above be such that $\delta_{\mathbf{y}}(B) < \varepsilon'$ for all $\mathbf{y} \in Q$.
In order to define L introduce the following partition of $[\mathbf{0}, \mathbf{1}[$ which is equal to Q up to a set of measure 0. Divide $[\mathbf{0}, \mathbf{1}[$ up into m^d equal subcubes Q_i, $i \in I := \{1, 2, \cdots, m^d\}$. Let $Q_i := [\mathbf{q}_i, \mathbf{q}_i + \frac{1}{m}\mathbf{1}[$ and the $\mathbf{q}_i$'s in Q be so that $\{q_i \,|\, i \in I\} = \{0, \frac{1}{m}, \frac{2}{m}, \cdots, \frac{m-1}{m}\}^d$.
Let

$$L := \bigcup_{i\in J} Q_i \qquad \text{with } J := \{i \in I \,|\, \mu([\mathbf{q}_i, \mathbf{1}[\cap B) \geq \varepsilon'\}.$$

By definition, L is a lower-layer: If $Q_i \subset L$ and Q_j is below Q_i in the sense that $\mathbf{q}_j \leq \mathbf{q}_i$ then $\mu([\mathbf{q}_j, \mathbf{1}[\cap B) \geq \mu([\mathbf{q}_i, \mathbf{1}[\cap B) \geq \varepsilon'$. So $Q_j \subset L$, too.
Now, we prove that $\mu(L \diamond B) < \varepsilon$ using $\mu(L \diamond B) = \mu(B \setminus L) + \mu(L \setminus B)$ and bounding the two terms separately.

$$\mu(B \setminus L) = \mu(B \cap L^c) = \mu(\bigcup_{i\in I\setminus J} Q_i \cap B) < m^d\varepsilon' = \varepsilon/3, \qquad (24)$$

since for $i \notin J$ we have $\mu(Q_i \cap B) \leq \mu([\mathbf{q}_i, \mathbf{1}[\cap B) < \varepsilon'$.
To bound the other part, first introduce

$$R := \{i \in J \,|\, \mu([\mathbf{q}_i + \frac{1}{m}\mathbf{1}, \mathbf{1}[\cap B) < \varepsilon'\}.$$

R is the set of all indexes i of cubes Q_i which lie in L at the 'boundary' of L and L^c. We can not get a good bound on $\mu((L \diamond B) \cap Q_i)$ for $i \in R$ but the idea is that this 'boundary' set is small enough (by our choice of m): We claim that $\#R \leq dm^{d-1}$.

The key observation is that if $q_i = q_j + \frac{k}{m}\mathbf{1}$ for some $k \geq 1$, then i and j can't both be in R, because $j \in R$ implies $\mu([\mathbf{q}_i, \mathbf{1}[\cap B) \leq \mu([\mathbf{q}_j + \frac{1}{m}\mathbf{1}, \mathbf{1}[\cap B) < \varepsilon'$, so i is not even in $J \supset R$.

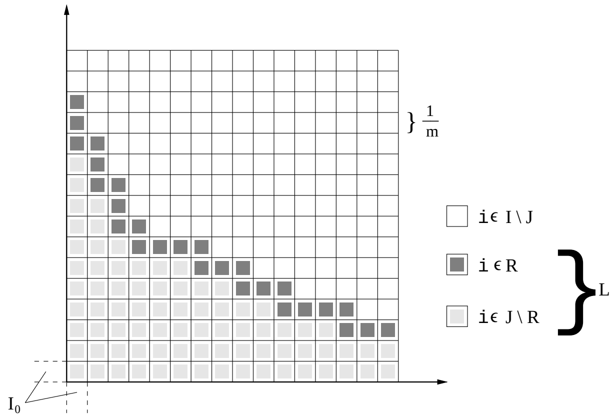


Figure 4: Construction of the lower-layer L

Let $I_0 := \{i \in I \,|\, \exists l \in \{1, 2, \cdots, d\} : q_i^{(l)} = 0\}$. For any $j \in I_0$ there is at most one $i \in R$ such that $\mathbf{q}_i = \mathbf{q}_j + \frac{k}{m}\mathbf{1}$ for some $k \geq 0$. But every $i \in R$ is covered that way. So $\#R \leq \#I_0$. And $\#I_0 \leq dm^{d-1}$, because for every $l \in \{1, 2, \ldots, d\}$ there are at most m^{d-1} $\boldsymbol{q}_i$'s with $\boldsymbol{q}_i^{(l)} = 0$. We get

$$\lambda(\bigcup_{i \in R} Q_i) \leq dm^{d-1}\frac{1}{m^d} = \frac{d}{m},$$

which, by the choice of m, implies

$$\mu(\bigcup_{i \in R} Q_i) \leq \varepsilon/3. \tag{25}$$

For $i \in J \setminus R$ we have $\mu([\mathbf{q}_i + \frac{1}{m}\mathbf{1}, \mathbf{1}[\cap B) \geq \varepsilon'$. But as $\delta_y(B) < \varepsilon'$ for all $\mathbf{y} \in B$, in particular for $\mathbf{y} = \mathbf{q}_i + \frac{1}{m}\mathbf{1}$ we conclude that $\mu([\mathbf{0}, \mathbf{q}_i + \frac{1}{m}\mathbf{1}] \cap B^c) < \varepsilon'$. As this set contains $Q_i \cap B^c$ we get

$$i \in J \setminus R \quad \Longrightarrow \quad \mu(Q_i \cap B^c) < \varepsilon'. \tag{26}$$

We are ready to bound

$$\begin{aligned}
\mu(L \setminus B) &= \mu\big(\bigcup_{i \in J} Q_i \cap B^c\big) \\
&\leq \mu(\bigcup_{i \in R} Q_i) + \sum_{i \in J \setminus R} \mu(Q_i \cap B^c) \\
&\leq \varepsilon/3 + m^d \varepsilon' \qquad\qquad \text{(because of (25) and (26))} \\
&= 2\varepsilon/3
\end{aligned}$$

(24) and (27) together imply $\mu(L \diamond B) \leq \varepsilon$. So $\mu(L) \leq \mu(B) + \varepsilon = \mu(A) + \varepsilon$. And (19) from the preceeding lemma implies $\mathbf{g}(L) \leq \mathbf{g}(B) + \varepsilon\,\mathbf{1} \leq \mathbf{g}(A) + \varepsilon\,\mathbf{1}$.

$\square$

A lower-layer A is in particular starlike with respect to the origin, i.e. $\mathbf{x} \in A \Rightarrow r\mathbf{x} \in A$ for $0 \leq r \leq 1$.
We can describe a starlike region A by a function in polar coordinates if $d > 1$. Define the following generalized polar coordinate transformation of the positive (and negative) orthant

$$\begin{aligned} \alpha : M \times \mathbb{R} &\rightarrow \mathbb{R}^d \\ (\boldsymbol{\varphi}, r) = (\varphi_1, \ldots, \varphi_{d-1}, r) &\mapsto \mathbf{x} = \alpha(\boldsymbol{\varphi}, r), \end{aligned}$$

where

$$M := [0, \frac{\pi}{2}]^{d-1}$$

and

$$\begin{aligned} x_1 &= r \cos \varphi_1 \\ x_2 &= r \sin \varphi_1 \cos \varphi_2 \\ x_3 &= r \sin \varphi_1 \sin \varphi_2 \cos \varphi_3 \\ &\vdots \\ x_{d-1} &= r \sin \varphi_1 \sin \varphi_2 \cdots \sin \varphi_{d-2} \cos \varphi_{d-1} \\ x_d &= r \sin \varphi_1 \sin \varphi_2 \quad \cdots \quad \sin \varphi_{d-1}. \end{aligned}$$

Then $\mathbf{x}$ ranges over $\mathbb{R}^d_+$ when r ranges over $\mathbb{R}_+$ and $\boldsymbol{\varphi}$ ranges over M.
We know (e.g. by [6]) that the functional determinant

$$\det D\alpha(\boldsymbol{\varphi}, r) = r^{d-1} \sin^{d-2} \varphi_1 \sin^{d-3} \varphi_2 \cdots \sin \varphi_{d-2}. \tag{27}$$

We will need this formula only to see that the functional determinant does not vanish for $r > 0$ and $\boldsymbol{\varphi} \in M^\circ$.
Now, define the function $R(\boldsymbol{\varphi}) = \sup\{r \mid \alpha(\boldsymbol{\varphi}, r) \in A\}$. Then $A \subset \{\alpha(\boldsymbol{\varphi}, r) \mid \boldsymbol{\varphi} \in M, 0 \leq r \leq R(\boldsymbol{\varphi})\}$ and the two sets differ only by a set of measure 0. We have a one-to-one correspondence up to sets of measure 0 between the starlike regions in Q and positive functions in polar coordinates. We will call R the function *describing* A.

In terms of R the measure $\mu(A)$ and the coordinates of $\mathbf{g}(A)$ are functionals $J(R)$ and

$G_j(R)$ $(1 \leq j \leq d)$, respectively.

$$\begin{aligned}
J(R) &:= \mu(A) \\
&= \int_{\mathbb{R}^d} 1\!\!1_A(\mathbf{x})\, f(\mathbf{x})\, d\mathbf{x} \\
&= \int_{M\times\mathbb{R}} 1\!\!1_A(\alpha(\mathbf{z}))\, f(\alpha(\mathbf{z}))\, |\det D\alpha(\mathbf{z})|\, d\mathbf{z} && \text{(transf. formula for Lebesgue integrals)} \\
&= \int_M \int_0^{R(\boldsymbol{\varphi})} f(\alpha(\boldsymbol{\varphi}, r))\, |\det D\alpha(\boldsymbol{\varphi}, r)|\, dr\, d\boldsymbol{\varphi} && \text{(Fubinis theorem)} \\
&= \int_M F(\boldsymbol{\varphi}, R(\boldsymbol{\varphi}))\, d\boldsymbol{\varphi},
\end{aligned}$$

where

$$F(\boldsymbol{\varphi}, R) := \int_0^R f(\alpha(\boldsymbol{\varphi}, r))\, |\det D\alpha(\boldsymbol{\varphi}, r)|\, dr. \tag{28}$$

Similarly,

$$\begin{aligned}
G_j(R) &:= g_j(A) = \int_{\mathbb{R}^d} x_j\, 1\!\!1_A(\mathbf{x})\, f(\mathbf{x})\, d\mathbf{x} \\
&= \int_M F_j(\boldsymbol{\varphi}, R(\boldsymbol{\varphi})) d\boldsymbol{\varphi},
\end{aligned}$$

where

$$F_j(\boldsymbol{\varphi}, R) := \int_0^R \alpha_j(\boldsymbol{\varphi}, r)\ f(\alpha(\boldsymbol{\varphi}, r))\, |\det D\alpha(\boldsymbol{\varphi}, r)|\, dr. \tag{29}$$

Lemma 2.15 (Lipschitz for lower-layers) If $A \subset Q$ is a lower-layer then the function $R : M \to \mathbb{R}_+$ describing A satisfies the Lipschitz condition with respect to the norm $\|\cdot\|_1$ on $M_\varepsilon := [\varepsilon, \frac{\pi}{2} - \varepsilon]^{d-1}$ with a constant $k = k(\varepsilon)$ that depends only on ε and not on R.
In other words

$$\boldsymbol{\beta}, \boldsymbol{\gamma} \in M_\varepsilon \quad \Longrightarrow \quad |R(\boldsymbol{\beta}) - R(\boldsymbol{\gamma})| \leq k\, \|\boldsymbol{\beta} - \boldsymbol{\gamma}\|_1$$

Proof. Let $1 > \varepsilon > 0$ be given and set

$$k := \frac{d}{(\frac{\varepsilon}{2})^d}.$$

Let $\boldsymbol{\beta}, \boldsymbol{\gamma} \in M_\varepsilon$ be arbitrary. Without loss of generality we can assume that $R(\boldsymbol{\gamma}) \geq R(\boldsymbol{\beta})$. Set $\mathbf{x}^\beta := \alpha(\boldsymbol{\beta}, R(\boldsymbol{\beta}))$ and $\mathbf{x}^\gamma := \alpha(\boldsymbol{\gamma}, R(\boldsymbol{\gamma}))$.
Because A is a lower-layer it is not possible that $\mathbf{x}^\gamma > \mathbf{x}^\beta$, because otherwise there would be a small $\delta > 0$ such that $(1-\delta)\mathbf{x}^\gamma > (1+\delta)\mathbf{x}^\beta$, too, but the first point is in A and the latter is not.
So there must be a coordinate i such that $x_i^\gamma \leq x_i^\beta$:

$$R(\boldsymbol{\gamma})w(\boldsymbol{\gamma}) \leq R(\boldsymbol{\beta})w(\boldsymbol{\beta}) \tag{30}$$

with $w(\boldsymbol{\varphi}) := \sin\varphi_1 \sin\varphi_2 \cdots \sin\varphi_{i-1} \cos^*\varphi_i$.
Letting $\Delta R := R(\boldsymbol{\gamma}) - R(\boldsymbol{\beta}) \geq 0$ we get from (30) that

$$\Delta R = R(\boldsymbol{\gamma}) - R(\boldsymbol{\beta}) = \frac{R(\boldsymbol{\gamma})w(\boldsymbol{\gamma}) - R(\boldsymbol{\beta})w(\boldsymbol{\gamma})}{w(\boldsymbol{\gamma})} \leq R(\boldsymbol{\beta})\,\frac{1}{w(\boldsymbol{\gamma})}\,(w(\boldsymbol{\beta}) - w(\boldsymbol{\gamma})), \tag{31}$$

in which be bound everything separately.
$R(\boldsymbol{\beta}) \leq d$, because $A \subset [0,1]^d$.
Since $\gamma_j \in [\varepsilon, \frac{\pi}{2} - \varepsilon]$, for $(j = 1, 2, \cdots, d)$, we have $\sin\gamma_j \geq \sin\varepsilon > \varepsilon/2$. We also get $\cos\gamma_j > \varepsilon/2$, using $\cos x = \sin(\frac{\pi}{2} - x)$. So $w(\boldsymbol{\gamma}) \geq (\frac{\varepsilon}{2})^d$.
Use the general formula $|\prod a_j - \prod b_j| \leq \sum |a_j - b_j|$, which holds for $|a_j|, |b_j| \leq 1$ (j ranges in the summation and the product over the same finite set.) to conclude

$$\begin{aligned} w(\boldsymbol{\beta}) - w(\boldsymbol{\gamma}) &\leq |\sin\beta_1 - \sin\gamma_1| + \cdots + |\sin\beta_{i-1} - \sin\gamma_{i-1}| + |\cos^*\beta_i - \cos^*\gamma_i| \\ &\leq |\beta_1 - \gamma_1| + \cdots + |\beta_i - \gamma_i| \\ &\leq \|\boldsymbol{\beta} - \boldsymbol{\gamma}\|_1. \end{aligned}$$

We conclude $\Delta R \leq k\|\boldsymbol{\beta} - \boldsymbol{\gamma}\|_1$, which was to be shown.

□

Lemma 2.16 (Existence of a solution)
Let μ be continuous with respect to the Lebesgue measure λ on Q and let f be a density of μ such that $\|f\|_\infty < \infty$.
Then (P_1) and (P_2) have a solution described by a continuous function in polar coordinates.

Remark: *The hypothesis $\|f\|_\infty < \infty$ will turn out unnecessary for (P_1) later (see the stronger theorem 2.17) but right now it facilitates the proof.*

Proof. Let $A_1, A_2, \ldots$ be a sequence of subsets of Q, which 'approaches optimality' in (P_1). That is $\mathbf{g}(A_n) \leq \boldsymbol{\rho}$ for all $n \in \mathbb{N}$ and $\mu(A_n) \to \text{Opt} := \sup\{\mu(A) \,|\, A \subset Q, \mathbf{g}(A) \leq \boldsymbol{\rho}\}$ as $n \to \infty$. And let L_n be a sequence of lower-layers such that $\mu(L_n) \geq \mu(A_n) - \frac{1}{n}$ and $\mathbf{g}(L_n) \leq \mathbf{g}(A_n) + \frac{1}{n}\,\mathbf{1}$, which exists by lemma 2.14, part (a).
Then

$$\liminf_n \mu(L_n) \geq \text{Opt} \qquad \text{and} \qquad \limsup_n\, g_j(L_n) \leq \rho_j \quad (j = 1, 2, \ldots, d) \tag{32}$$

*In case $i = d$ this cosine must be a sine.

In the same way we get – using part (b) of lemma 2.14 – a sequence of lower-layers L'_n such that

$$\liminf_n \mu(L'_n) \geq \mathrm{Opt}' \qquad \text{and} \qquad \limsup_n \, c_j(L'_n) \leq \tau_j \quad (j = 1, 2, \ldots, d) \tag{33}$$

with $\mathrm{Opt}' := \sup\,\{\mu(A)\,|\,A \subset Q, \mathbf{c}(A) \leq \boldsymbol{\tau}\}$. This is possible since $\mathrm{Opt}' > 0$, because of $\boldsymbol{\tau} > \mathbf{0}$ and (10).

Let $R_n : M \to \mathbb{R}_+$ be the function in polar coordinates describing L_n.
If $\varepsilon > 0$ is given and M_ε is like in lemma 2.15 then R_n satisfies the Lipschitz condition with some constant k, which doesn't depend on R_n by lemma 2.15.
This means that $\{R_1, R_2, \ldots\}$ is equicontinuous on M_ε: For given $\varepsilon' > 0$ there is a $\delta > 0$, namely $\delta := \varepsilon'/k$ such that for $\boldsymbol{\varphi_1}, \boldsymbol{\varphi_2} \in M_\varepsilon$ $\|\boldsymbol{\varphi_1} - \boldsymbol{\varphi_2}\|_1 < \delta$ implies $|R_n(\boldsymbol{\varphi_1}) - R_n(\boldsymbol{\varphi_2})| \leq k\|\boldsymbol{\varphi_1} - \boldsymbol{\varphi_2}\|_1 \leq k\delta = \varepsilon'$ for all $n \in \mathbb{N}$.
Also every R_n is bounded from below by 0 and from above by d.
We can apply the theorem of Arzelà-Ascoli (see theorem 4.10) to get that R_n contains a subsequence that converges uniformly on M_ε. As $\varepsilon > 0$ was arbitrary, we can choose $\varepsilon = \frac{1}{k}$. Let $(R_1^1, R_2^1, R_3^1, \ldots)$ be a subsequence of $(R_1, R_2, R_3 \ldots)$ that converges uniformly on M_1. And if $(R_n^k)_{n\in\mathbb{N}}$ is already defined, let $(R_n^{k+1})_{n\in\mathbb{N}}$ be a subsequence of $(R_n^k)_{n\in\mathbb{N}}$ that converges uniformly on $M_{1/(k+1)}$. Then the diagonal sequence $(R_n^n)_{n\in\mathbb{N}}$ converges pointwise on $]0, \frac{\pi}{2}[^{d-1} = M^\circ$ to a function R. Also $(R_n^n)_{n\in\mathbb{N}}$ convergences uniformly on every M_ε and the R_n^n are continuous on M°. Thus R is continuous on M° and can be extended to a continuous function on M. For simplicity, write $\hat{R}_n := R_n^n$, $\hat{L}_n$ for the region described by $\hat{R}_n$ and L for the region described by R.
As $\hat{R}_n \to R$ pointwise on M° , we have $\hat{R}_n \to R$ in $L_1(M)$, too, by the theorem of Lebesgue ($|\hat{R}_n|$ is dominated by the constant d).
The latter convergence implies that

$$\mu(\hat{L}_n) \to \mu(L) \qquad \text{and} \qquad \mathbf{g}(\hat{L}_n) \to \mathbf{g}(L), \tag{34}$$

because

$$\begin{aligned}
|\mu(\hat{L}_n) - \mu(L)| &= |J(\hat{R}_n - J(R)| \\
&= \Big|\int_M F(\boldsymbol{\varphi}, \hat{R}_n(\boldsymbol{\varphi})) - F(\boldsymbol{\varphi}, R(\boldsymbol{\varphi}))\, d\boldsymbol{\varphi}\Big| \qquad \text{(with } F \text{ defined as in (28))} & (35)\\
&\leq \int_M \Big| \int_{R(\boldsymbol{\varphi})}^{\hat{R}_n(\boldsymbol{\varphi})} f(\alpha(\boldsymbol{\varphi}, r))\, |\det D\alpha(\boldsymbol{\varphi}, r)|\, dr \Big|\, d\boldsymbol{\varphi} & (36)\\
&\leq \int_M |\hat{R}_n(\boldsymbol{\varphi}) - R(\boldsymbol{\varphi})|\, \|f\|_\infty\, d^d\, d\boldsymbol{\varphi} & (37)\\
&= \|f\|_\infty\, d^d\, \|\hat{R}_n - R\|_1 \to 0.
\end{aligned}$$

The proof for $\mathbf{g}(\hat{L}_n) \to \mathbf{g}(L)$ is the same with only a little change: We have $|g_j(\hat{L}_n) - g_j(L)| = |G_j(\hat{R}_n) - G(R)|$. Then in (35) it must be F_j instead of F, which is defined

in (29). And in (36) there also is an additional – third – factor $\alpha_j(\boldsymbol{\varphi}, r)$ in the inner integrand, whose absolute value is bounded from above by 1 in (37).
(34) enables us to conclude that L is solution to (P_1): As $(\hat{L}_n)$ is a subsequence of (L_n) we can conclude by (32) that

$$\mu(L) \geq \text{Opt} \quad \text{and} \quad \mathbf{g}(L) \leq \boldsymbol{\rho}.$$

With the same argument we also have a subsequence $(\hat{L}'_n)$ of (L'_n) and a set L' described by a continuous function R' such that

$$\mu(\hat{L}'_n) \to \mu(L') \qquad \text{and} \qquad \mathbf{g}(\hat{L}'_n) \to \mathbf{g}(L'). \tag{38}$$

Again (33) implies that $\mu(L') \geq \text{Opt}'$. And as $\text{Opt}' > 0$ – and therefore $\mu(L') > 0$, too – this implies

$$\mathbf{c}(\hat{L}'_n) = \frac{\mathbf{g}(\hat{L}'_n)}{\mu(\hat{L}'_n)} \to \frac{\mathbf{g}(L')}{\mu(L')} = \mathbf{c}(L').$$

So, by (33), $\mathbf{c}(L') \leq \boldsymbol{\tau}$, which proves that L' is solution to (P_2).

□

2.3.4 Shape of the solution

We will now show that an optimal region $\hat{A}$ is a simplicial section up to sets of measure 0.

Theorem 2.17 Given a probability measure μ on Q with a density f with respect to Lebesgue measure and a $\boldsymbol{\rho} > \mathbf{0}$ there is a simplicial section $\hat{A} = \{\mathbf{x} \in Q \mid \langle \mathbf{x}, \boldsymbol{\theta} \rangle \leq 1\}$ which soves (P_1). That is, it maximizes $\mu(A)$ over all $A \subset Q$ which have $\mathbf{g}(A) \leq \boldsymbol{\rho}$.
Furthermore, for all $i \in \{1, \ldots, d\}$ such that the i-th constraint is inactive, i.e. $g_i(\hat{A}) < \rho_i$, we have $\theta_i = 0$.
The optimal region is unique up to sets of measure 0.

Proof.
Part 1. First assume that the density f is continuous on $\mathbb{R}^d$, $f(\mathbf{x}) > 0$ for $\mathbf{x}$ in Q° and $f(\mathbf{x}) = 0$ for $\mathbf{x} \notin Q$. We will later approximate the general density f of μ by continuous densities like above.
Let $\hat{A} \subset Q$ be a starlike *optimal* region given by the continuous function $\hat{R}(\boldsymbol{\varphi})$. The existence is ensured by lemma 2.16 as $\|f\|_\infty < \infty$ in this case. And let A be any (measurable) starlike region, given by the function R.
We want to apply the generalized Kuhn-Tucker theorem (see theorem (4.4)). Let H be the vector space of all bounded measurable functions on M.
The optimal solution $\hat{R}$ miminizes $-J(R)$ (i.e maximizes $J(R)$) over all $R \in H$ satisfying the constraint

$$\mathbf{G}(R) - \boldsymbol{\rho} \leq \mathbf{0}.$$

The fact that H contains functions R which attain negative values or don't describe a subset of Q does not bring complications. If R is such that for some $\boldsymbol{\varphi} \in M$ $\alpha(\boldsymbol{\varphi}, R(\boldsymbol{\varphi})) \notin Q$ we can define $R^*(\boldsymbol{\varphi})$ to be 0 if $R(\boldsymbol{\varphi})$ is negative and maximal so that $\alpha(\boldsymbol{\varphi}, R^*(\boldsymbol{\varphi})) \in Q$ if $R(\boldsymbol{\varphi})$ was too large. Recall that we have $f(\mathbf{x}) = 0$ for $x \notin Q$, so $J(R) = J(R^*)$ and $\mathbf{G}(R) = \mathbf{G}(R^*)$.
We have to show that J and G are Gateaux differentiable functionals (see definition 4.3) on H and that the variations are linear in their increments.
For any $R, h \in H$ the Gateaux-variation of J at R with increment h is (if it exists)

$$\begin{aligned} \delta J(R, h) &= \frac{d}{d\varepsilon} J(R + \varepsilon h)\Big|_{\varepsilon=0} \\ &= \frac{d}{d\varepsilon} \int_M F(\boldsymbol{\varphi}, R + \varepsilon h)\, d\boldsymbol{\varphi}\Big|_{\varepsilon=0}. \end{aligned} \tag{39}$$

The integrand $F(\boldsymbol{\varphi}, R + \varepsilon h)$ is differentiable with respect to ε: For every $\boldsymbol{\varphi}$ the integrand in the definition of F

$$f(\alpha(\boldsymbol{\varphi}, r))\, |\det D\alpha(\boldsymbol{\varphi}, r)|$$

is continuous in r, since α, f and $\det D\alpha(\boldsymbol{\varphi}, r)$ are continuous. And therefore we get

$$\begin{aligned} \frac{d}{d\varepsilon} F(\boldsymbol{\varphi}, R + \varepsilon h) &= \frac{\partial F}{\partial R}(\boldsymbol{\varphi}, R + \varepsilon h)\, h(\boldsymbol{\varphi}) \\ &= f(\alpha(\boldsymbol{\varphi}, R + \varepsilon h))\, |\det D\alpha(\boldsymbol{\varphi}, R + \varepsilon h)|\, h(\boldsymbol{\varphi}), \end{aligned}$$

which is bounded in $\boldsymbol{\varphi}$ and ε. Therefore we can differentiate in (39) under the integral sign and get

$$\delta J(R,h) = \int\limits_M f(\alpha(\boldsymbol{\varphi},R)) \, |\det D\alpha(\boldsymbol{\varphi},R)| \, h(\boldsymbol{\varphi}) \, d\boldsymbol{\varphi}. \tag{40}$$

It was essential here that f is continuous.
In the same manner we get the Gateaux-variations of G_j $(j = 1,\dots,d)$:

$$\delta G_j(R,h) = \int\limits_M \alpha_j(\boldsymbol{\varphi},R) f(\alpha(\boldsymbol{\varphi},R)) \, |\det D\alpha(\boldsymbol{\varphi},R)| \, h(\boldsymbol{\varphi}) \, d\boldsymbol{\varphi}. \tag{41}$$

Since limits in $\mathbb{R}^d$ are defined componentwise, we get

$$\delta \mathbf{G}(R,h) = (\delta G_1(R,h),\dots,\delta G_d(R,h)).$$

And we see that both $\delta J(R,h)$ and $\delta\mathbf{G}(R,h)$ are linear in the increment h.
The last thing we need to show to be able to apply the theorem of Kuhn-Tucker is that $\hat{R}$ is a regular point of the inequality $\mathbf{G}(R) - \boldsymbol{\rho} \leq \mathbf{0}$.
Since $\mathbf{G}(\hat{R}) \leq \boldsymbol{\rho}$ it is sufficient to give an $h \in H$ such that

$$\delta\mathbf{G}(\hat{R},h) < \mathbf{0}.$$

Simply choose $h(\boldsymbol{\varphi}) = -1$ for all $\boldsymbol{\varphi} \in M$. Then formula (41) gives

$$\delta G_j(\hat{R},h) = -\int\limits_M \alpha_j(\boldsymbol{\varphi},\hat{R}) f(\alpha(\boldsymbol{\varphi},\hat{R})) \, |\det D\alpha(\boldsymbol{\varphi},\hat{R})| \, d\boldsymbol{\varphi}.$$

As $\hat{R}$ is continuous, either $\hat{A} = Q$, $\hat{A} = \{\mathbf{0}\}$ or $\alpha(\boldsymbol{\varphi},\hat{R}) \in Q^\circ$ for $\boldsymbol{\varphi}$ in a set of positive Lebesgue measure. If $\hat{A} = Q$, we are done since Q is a simplicial section with $\boldsymbol{\theta} = \mathbf{0}$. $\hat{A} = \{\mathbf{0}\}$ is impossible: $\mu(\hat{A})$ must be positive because $\boldsymbol{\rho} > \mathbf{0}$. In the other case we get $\delta\mathbf{G}(\hat{R},h) < 0$, because then all three factors of the integrand are positive for $\boldsymbol{\varphi}$ in a set of positive measure.
Now, by theorem (4.4) we have a $\boldsymbol{\theta} = (\theta_1,\dots,\theta_d) \geq \mathbf{0}$ such that the Lagrangian

$$L(R) := \langle \mathbf{G}(R) - \boldsymbol{\rho}, \boldsymbol{\theta} \rangle - J(R)$$

is stationary at $\hat{R}$ and

$$\left\langle \mathbf{G}(\hat{R}) - \boldsymbol{\rho}, \boldsymbol{\theta} \right\rangle = 0. \tag{42}$$

As $\hat{R}$ satisfies the constraint $\mathbf{G}(\hat{R}) - \boldsymbol{\rho} \leq \mathbf{0}$ and as $\boldsymbol{\theta} \geq 0$, we can conclude by (42) that

$$\theta_i = 0 \quad \text{if} \quad g_i(\hat{A}) = G_i(\hat{R}) < \rho_i,$$

as desired.
It remains to prove that the simplicial section $\{\mathbf{x} \in Q \mid \langle x, \boldsymbol{\theta} \rangle \leq 1\}$ is an optimal region.
By definition of a stationary point we have for every $h \in H$

$$\begin{aligned} 0 &= \delta L(\hat{R}, h) \\ &= \left\langle \delta \mathbf{G}(\hat{R}, h), \boldsymbol{\theta} \right\rangle - \delta \mathbf{J}(\hat{R}, h) \\ &= \int_M \underbrace{(\left\langle \alpha(\boldsymbol{\varphi}, \hat{R}), \boldsymbol{\theta} \right\rangle - 1)\, f(\alpha(\boldsymbol{\varphi}, \hat{R}))\, |\det D\alpha(\boldsymbol{\varphi}, \hat{R})|}_{=:l(\boldsymbol{\varphi})}\, h(\boldsymbol{\varphi})\, d\boldsymbol{\varphi}. \end{aligned} \tag{43}$$

Since h is arbitrary, we can use

$$h(\boldsymbol{\varphi}) := \begin{cases} 1 & , \text{ if } l(\boldsymbol{\varphi}) \geq 0 \\ -1 & , \text{ if } l(\boldsymbol{\varphi}) < 0 \end{cases}$$

Then $l\, h$ is nonnegative and vanishes exactly when l vanishes. With (43) we conclude that λ-almost surely $l(\boldsymbol{\varphi}) = 0$.
For $\alpha(\boldsymbol{\varphi}, \hat{R}) \in Q^\circ$, $f(\alpha(\boldsymbol{\varphi}, \hat{R})) > 0$ by hypothesis and $\det D\alpha(\boldsymbol{\varphi}, \hat{R}) \neq 0$. So, either $\alpha(\boldsymbol{\varphi}, \hat{R}) \in \partial Q$ or

$$\left\langle \alpha(\boldsymbol{\varphi}, \hat{R}), \boldsymbol{\theta} \right\rangle = 1$$

Let K be the hyperplane $\{\mathbf{x} \mid \langle \mathbf{x}, \boldsymbol{\theta} \rangle = 1\}$. Then $\partial\hat{A} = \{\alpha(\boldsymbol{\varphi}, \hat{R}(\boldsymbol{\varphi})) \mid \boldsymbol{\varphi} \in M\} \subset (K \cap Q) \cup \partial Q$. So as $\hat{R}$ is continuous we must have that $\hat{A} = \{\mathbf{x} \in Q \mid \langle \mathbf{x}, \boldsymbol{\theta} \rangle \leq 1\}$ or $\hat{A} = Q$ and that was correct, too.

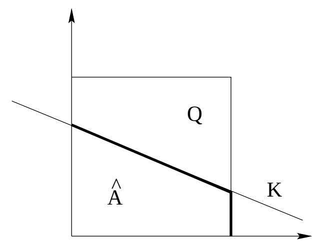


Figure 5: The simplicial section $\hat{A}$

Part 2. Now, drop the additional assumptions made in part 1 about the density f of μ. Let μ have the density f, which still is defined on $\mathbb{R}^d$ (that is $f(\mathbf{x}) = 0$ a.s. for $\mathbf{x} \notin Q$), but does not need to be positive on Q° or continuous.
Let $f_1, f_2, \ldots$ be a sequence of probability densities such that f_n fulfills for every n the hypothesis of part 1 and the sequence (f_n) converges to f with respect to the L_1-norm, i.e.

- f_n is continuous
- $f_n(\mathbf{x}) > 0$ for $\mathbf{x} \in Q^\circ$, $f_n(\mathbf{x}) = 0$ for $\mathbf{x} \notin Q$
- $\int |f - f_n|\, d\lambda \to 0$

We will now show that such an (f_n) indeed exists. We will use that the set of continuous functions is dense with respect to the L_1-norm in the set of Lebesgue-integrable functions. As $f \geq 0$ there is a continuous function $g : \mathbb{R}^d \to \mathbb{R}$, $g \geq 0$, such that

$$\int |g - f| \, d\lambda < \varepsilon \tag{44}$$

for $\varepsilon := 1/n$.
This g doesn't need to be positive on Q°, nor vanish on $\mathbb{R}^d \setminus Q$ nor be a probability density (i.e. $\int g \, d\lambda = 1$). But we can achieve these properties by making three 'small changes' to g.
First, let $g' := g + \varepsilon \mathbb{1}_Q$, so g' is positive on Q° and still continuous and bounded on Q°. g' is integrable as g is integrable (this is because of (44)), so we can choose a closed subsquare B of Q° such that

$$\int\limits_{Q \setminus B} f \, d\lambda < \varepsilon \quad \text{and} \quad \int\limits_{Q \setminus B} g' \, d\lambda < \varepsilon. \tag{45}$$

Use the strong version of the lemma of Urysohn to get a continuous function $h : \mathbb{R}^d \to [0,1]$ with $h(\mathbf{x}) = 0$ for $\mathbf{x} \in \mathbb{R}^d \setminus Q^\circ$, $h(\mathbf{x}) = 1$ for $\mathbf{x} \in B$ and $h(\mathbf{x}) \in (0,1)$, otherwise.
Now, define $g'' := g'h$. g'' is continuous on $\mathbb{R}^d$ again, since g' was continuous everywhere except possibly on ∂Q. But g'' is continuous on ∂Q because h vanishes on ∂Q and g' is bounded on Q°. g'' still is positive on Q° and g'' vanishes on $\mathbb{R}^d \setminus Q$ so it has all the desired properties except that $m := \int g'' \, d\lambda \neq 1$ maybe.
Finally, define $f_n(\mathbf{x}) := \frac{1}{m} g''(\mathbf{x})$. Then f_n fulfills what we have claimed. It remains to show that (f_n) converges to f in L_1.
We will show in three steps that g', g'', and f_n are close to f in L_1.
Firstly,

$$\int |g' - f| \, d\lambda \leq \int |g - f| \, d\lambda + \int |g - g'| \, d\lambda \leq \varepsilon + \varepsilon = 2\varepsilon. \tag{46}$$

Secondly,

$$\begin{aligned} \int |g'' - f| \, d\lambda &= \int\limits_B |g'' - f| \, d\lambda + \int\limits_{Q \setminus B} |g'' - f| \, d\lambda + \int\limits_{\mathbb{R}^d \setminus Q} |g'' - f| \, d\lambda \\ &\leq \int\limits_B |g' - f| \, d\lambda + \int\limits_{Q \setminus B} \underbrace{h \, g'}_{\leq g'} + f \, d\lambda + 0, \end{aligned}$$

since $g'' = g'$ on B and $g'' = f = 0$ a.s. on $\mathbb{R}^d \setminus Q$. By (46) and (45)

$$\int |g'' - f| \, d\lambda \leq 2\varepsilon + \varepsilon + \varepsilon = 4\varepsilon. \tag{47}$$

This also gives us $m = \int g'' \, d\lambda \to 1$ as $n \to \infty$, because of $\int f \, d\lambda = 1$.
Thirdly,

$$\begin{aligned}\int |f_n - f| \, d\lambda &= \int |\frac{1}{m} g'' - f| \, d\lambda \\ &\le \int |g'' - f| + |1 - \frac{1}{m}| g'' \, d\lambda \\ &\le 4\varepsilon + |m - 1| \to 0.\end{aligned}$$

Now define $\mu_n := f_n \lambda$ for $n = 1, 2, \ldots$. So μ_n is a probability measure on Q with density f_n and part 1 is applicable to μ_n. Also, for brevity, define $\mathbf{g}^n(A) := \int_A \mathbf{x} \, d\mu_n(\mathbf{x})$ and $\mathbf{c}^n(A) := \mathbf{g}^n(A)/\mu_n(A)$ in analogy to $\mathbf{g}(A)$ and $\mathbf{c}(A)$.
Since $f_n \to f$ in L_1 we now have

$$\mu_n(A) \to \mu(A) \qquad \text{and} \qquad \mathbf{g}^n(A) \to \mathbf{g}(A) \qquad (A \subset Q). \tag{48}$$

Let $\Delta_n = \{\mathbf{x} \,|\, \langle x, \boldsymbol{\theta}^{(n)} \rangle \le 1\}$ be the simplicial section from part 1. I.e., it is an optimal region for the measure μ_n and $A = \Delta_n$ maximizes $\mu_n(A)$ over $\mathbf{g}^n(A) \le \boldsymbol{\rho}$.
The idea is that a subsequence of (Δ_n) converges to a simplicial section Δ in such a sense that we can conclude that this Δ is optimal for the probability measure μ.
We will show that a subsequence of $\boldsymbol{\theta}^{(n)}$ converges in $\mathbb{R}^d$ with the usual topology to a vector $\boldsymbol{\theta}$.
First observe that, as $\boldsymbol{\rho} > \mathbf{0}$, we can show that the $\boldsymbol{\theta}^{(n)}$'s are bounded. Consider the i-th coordinate of $\boldsymbol{\theta}^{(n)}$. By part 1, either $\theta_i^{(n)} = 0$ or $\rho_i = g_i^n(\Delta_n) \le c_i^n(\Delta_n) \le$
$\max\{x_i \,|\mathbf{x} \in \Delta_n\} \le 1/\theta_i^{(n)}$. So $\theta_i^{(n)} \le 1/\rho_i$. With $M := \max_i 1/\rho_i$ we have $\boldsymbol{\theta}^{(n)} \in [0, M]^d$.
Since this set is compact, there is a convergent subsequence of $(\boldsymbol{\theta}^{(n)})$. Let $\boldsymbol{\theta} \in \mathbb{R}_+^d$ be the limit. For simplicity with the notations assume that this subsequence was the sequence $(\boldsymbol{\theta}^{(n)})$ itself. Define

$$\Delta := \{\mathbf{x} \in Q \mid \langle \mathbf{x}, \boldsymbol{\theta} \rangle \le 1\}. \tag{49}$$

We will now show that $\lambda(\Delta_n \diamond \Delta)$ (the volume of the symmetric difference of the sets Δ_n and Δ) converges to 0.
We can assume that $\boldsymbol{\theta} \ne \mathbf{0}$. Otherwise, $\Delta = Q$ and, as $\boldsymbol{\theta}^{(n)} \to \boldsymbol{\theta}$, for n large enough $\theta_1^{(n)} + \theta_2^{(n)} + \cdots + \theta_d^{(n)} \le 1$. So that $\mathbf{1} \in \Delta_n$, since $\langle \mathbf{1}, \boldsymbol{\theta}^{(n)} \rangle = \theta_1^{(n)} + \theta_2^{(n)} + \cdots + \theta_d^{(n)} \le 1$. But then $\Delta_n = Q$, too, and $\Delta \diamond \Delta_n = \emptyset$ so we are done.
Now, let $\|\boldsymbol{\theta}\|$, the euclidian norm of $\boldsymbol{\theta}$, be positive. By Cauchy-Schwarz's inequality $|\langle \mathbf{x}, \boldsymbol{\theta} - \boldsymbol{\theta}^{(n)} \rangle| \le \|\mathbf{x}\| \|\theta - \boldsymbol{\theta}^{(n)}\| \to 0$ uniformly for $\mathbf{x} \in Q$. Thus for any $\varepsilon > 0$ there is an n_0 such that for $n > n_0$

$$| \langle \mathbf{x}, \theta \rangle - \langle \mathbf{x}, \boldsymbol{\theta}^{(n)} \rangle | < \varepsilon \qquad (x \in Q). \tag{50}$$

Now, suppose $\mathbf{x} \in \Delta \diamond \Delta_n$. Then $\langle \mathbf{x}, \boldsymbol{\theta} \rangle \le 1$ and $\langle \mathbf{x}, \boldsymbol{\theta}^{(n)} \rangle > 1$ or the other way around. Hence, because of (50), we get

$$\begin{aligned}\Delta \diamond \Delta_n &\subset \{\mathbf{x} \in Q \,|\, \langle \mathbf{x}, \theta \rangle \in (1 - \varepsilon, 1 + \varepsilon)\} \\ &= \{\mathbf{x} \in Q \,|\, \left\langle \mathbf{x}, \frac{\theta}{\|\boldsymbol{\theta}\|} \right\rangle \in (\frac{1 - \varepsilon}{\|\boldsymbol{\theta}\|}, \frac{1 + \varepsilon}{\|\boldsymbol{\theta}\|})\}.\end{aligned}$$

This is the intersection of Q with the set of all points $\mathbf{x}$ which lie between two parallel hyperplanes with Euclidean distance $2\varepsilon/\|\boldsymbol{\theta}\|$. Because Q is bounded (its diameter is $\sqrt{d}$) we get

$$\limsup_n \lambda(\Delta \diamond \Delta_n) \leq \frac{2\varepsilon}{\|\boldsymbol{\theta}\|}\sqrt{d}.$$

Letting $\varepsilon \to 0$ this yields

$$\lim_{n\to\infty} \lambda(\Delta \diamond \Delta_n) = 0.$$

Since μ is continuous with respect to the Lebesgue measure we also get

$$\lim_{n\to\infty} \mu(\Delta \diamond \Delta_n) = 0. \tag{51}$$

which implies by lemma 2.13 that

$$\mathbf{g}(\Delta_n) \to \mathbf{g}(\Delta). \tag{52}$$

Finally we show that $A = \Delta$ maximizes $\mu(A)$ over $\mathbf{g}(A) \leq \boldsymbol{\rho}$ and $g_i(\Delta) < \rho_i \Rightarrow \theta_i = 0$. Suppose for the sake of contradiction there was an $A \subset Q$ with $\mu(A) > \mu(\Delta)$ and $\mathbf{g}(A) \leq \boldsymbol{\rho}$. Then there also would be an A' with $\mu(A) > \mu(\Delta)$ and $\mathbf{g}(A) < \boldsymbol{\rho}$. This is so because we simply could take away from A a small subset $B \subset A$: $0 < \mu(B) < \mu(A) - \mu(\Delta)$. Then $A' := A \setminus B$ would have $g_i(A') = g_i(A) - \int_B x_i \, d\mu(\mathbf{x}) < g_i(A) \leq \rho_i$ and $\mu(A') > \mu(A)$.
We want to show that then some Δ_n could not have been optimal. First observe that

$$\begin{aligned} |\mu(\Delta) - \mu_n(\Delta_n)| &\leq |\mu(\Delta) - \mu(\Delta_n)| + |\mu(\Delta_n) - \mu_n(\Delta_n)| \\ &\leq \mu(\Delta \diamond \Delta_n) + \int |f - f_n| \, d\lambda \\ &\to 0 \end{aligned} \tag{53}$$

where we used (51) and made use of the choice of f_n.
Similarly,

$$\begin{aligned} \|\mathbf{g}(\Delta) - \mathbf{g}^n(\Delta_n)\| &\leq \|\mathbf{g}(\Delta) - \mathbf{g}(\Delta_n)\| + \|\mathbf{g}(\Delta_n) - \mathbf{g}^n(\Delta_n)\| \\ &\leq \|\mathbf{g}(\Delta) - \mathbf{g}(\Delta_n)\| + \int_{\Delta_n} \|(f(\mathbf{x}) - f_n(\mathbf{x}))\mathbf{x}\| \, d\mathbf{x} \\ &\leq \|\mathbf{g}(\Delta) - \mathbf{g}(\Delta_n)\| + \sqrt{d}\int |f - f_n| \, d\lambda \\ &\to 0 \end{aligned} \tag{54}$$

because of (52).

Now, define $\varepsilon := \mu(A') - \mu(\Delta)$ and choose n so large that

(1) $\mathbf{g}^n(A') < \boldsymbol{\rho}$ (possible, since $\mathbf{g}^n(A') \to \mathbf{g}(A') < \boldsymbol{\rho}$ and by (48))

(2) $\mu_n(A') > \mu(A') - \frac{\varepsilon}{2}$ (possible by (48))

(3) $\mu_n(\Delta_n) < \mu(\Delta) + \frac{\varepsilon}{2}$. (possible by (53))

Then

$$\mu_n(A') \overset{(2)}{>} \mu(A') - \frac{\varepsilon}{2} = \mu(\Delta) + \frac{\varepsilon}{2} \overset{(3)}{>} \mu_n(\Delta_n),$$

which contradicts together with (1) the assumption that Δ_n was an optimal region for the measure μ_n.
And suppose $g_i(\Delta) < \rho_i$ then $g_i^{(n)}(\Delta_n) < \rho_i$ for n large enough, because of (54). This implies $\theta_i^{(n)} = 0$ by part 1. So $\theta_i = \lim_{n\to\infty} \theta_i^{(n)} = 0$, too.

It remains to show that the optimal region is unique up to sets of measure 0. Let Δ be as in (49) an optimal region and let B be any other optimal region. We will show that $\mu(\Delta \diamond B) = 0$.
Suppose $\mu(\Delta \diamond B) > 0$. Then $\mu(B \setminus \Delta) = \mu(\Delta \setminus B) =: m > 0$, as $\mu(\Delta) = \mu(B)$.
We have

$$\langle \mathbf{g}(B), \boldsymbol{\theta}\rangle - \langle \mathbf{g}(\Delta), \boldsymbol{\theta}\rangle = \int\limits_{B\setminus\Delta} \langle \mathbf{x}, \boldsymbol{\theta}\rangle \, d\mu(\mathbf{x}) - \int\limits_{\Delta\setminus B} \langle \mathbf{x}, \boldsymbol{\theta}\rangle \, d\mu(\mathbf{x}) \tag{55}$$

But for $\mathbf{x} \in B \setminus \Delta$ we have $\langle \mathbf{x}, \boldsymbol{\theta}\rangle > 1$ and for $\mathbf{x} \in \Delta \setminus B$ we have $\langle \mathbf{x}, \boldsymbol{\theta}\rangle \leq 1$. So the right integral is at most m. Plugging this into (55) yields

$$\langle \mathbf{g}(B), \boldsymbol{\theta}\rangle - \langle \mathbf{g}(\Delta), \boldsymbol{\theta}\rangle \geq \int\limits_{B\setminus\Delta} \underbrace{\langle \mathbf{x}, \boldsymbol{\theta}\rangle - 1}_{>0} \, d\mu(\mathbf{x}) > 0$$

as $\mu(B \setminus \Delta) > 0$. We get

$$\langle \mathbf{g}(B), \boldsymbol{\theta}\rangle > \langle \mathbf{g}(\Delta), \boldsymbol{\theta}\rangle = \langle \boldsymbol{\rho}, \boldsymbol{\theta}\rangle \tag{56}$$

where the equation on the right holds because $\theta_i = 0$ if $g_i(\Delta) \neq \rho_i$. But (56) implies that $g_i(B) > \rho_i$ for some $i \in \{1, \dots, d\}$, which is a contradiction. So $\mu(\Delta \diamond B) = 0$.

□

Corollary 2.18 Given a probability measure μ on Q with a density f such that $\|f\|_\infty < \infty$ and given a $\boldsymbol{\tau} > \mathbf{0}$ there is a simplicial section $\hat{A} = \{\mathbf{x} \in Q \mid \langle \mathbf{x}, \boldsymbol{\theta}\rangle \leq 1\}$ which solves (P_2). That is, it maximizes $\mu(A)$ amongst all $A \subset Q$ which have $\mathbf{c}(A) \leq \boldsymbol{\tau}$.
For all i such that the i-th constraint is inactive, i.e. $c_i(\hat{A}) < \tau_i$, we have $\theta_i = 0$.
Furthermore, the optimal region is unique up to sets of measure 0.

Proof. Let A be a solution to (P_2), i.e. $\mathbf{c}(A) \leq \boldsymbol{\tau}$ and whenever $\mathbf{c}(A') \leq \boldsymbol{\tau}$ we have $\mu(A') \leq \mu(A)$. A exists because of lemma 2.16. We have $\mu(A) > 0$ because $\boldsymbol{\tau} > \mathbf{0}$. Define $\boldsymbol{\rho} > \mathbf{0}$ by

$$\boldsymbol{\rho} := \mu(A)\boldsymbol{\tau} \in Q.$$

We will show that A also maximizes $\mu(A)$ under the restriction $\mathbf{g}(A) \leq \boldsymbol{\rho}$.
First note that A satisfies this restriction since $\mathbf{g}(A) = \mu(A)\mathbf{c}(A) \leq \mu(A)\boldsymbol{\tau} = \boldsymbol{\rho}$.
Now, let A' be any set such that $\mathbf{g}(A') \leq \boldsymbol{\rho}$. Then we have

$$\begin{aligned} \mu(A')\mathbf{c}(A') &= \mathbf{g}(A') \\ &\leq \boldsymbol{\rho} \\ \mu(A')\mathbf{c}(A') &= \mu(A)\boldsymbol{\tau} \end{aligned} \tag{57}$$

Now, if $\mathbf{c}(A') \leq \boldsymbol{\tau}$, we have $\mu(A') \leq \mu(A)$ since $\mu(A)$ was maximal under that condition. And if $\mathbf{c}(A') \not\leq \boldsymbol{\tau}$ there is at least one coordinate i such that $c_i(A') > \tau_i$. But then the i-th coordinate of inequality (57) tells us that $\mu(A') \leq \mu(A)$.
Since A' was arbitrary, A also maximizes $\mu(A)$ under the restriction $\mathbf{g}(A) \leq \boldsymbol{\rho}$.
Because the solution to that problem is unique and a lower-layer, by theorem 2.17, we get that A is a simplicial section $\hat{A} = \{\mathbf{x} \in Q \mid \langle \mathbf{x}, \boldsymbol{\theta} \rangle \leq 1\}$ up to sets of measure 0. And since $\mu(A) = \mu(\hat{A})$ we have $\theta_i = 0$ for each i such that $g_i(\hat{A}) < \rho_i = \mu(\hat{A})\tau_i$. Or equivalently,

$$\theta_i = 0 \qquad \text{if} \quad c_i(\hat{A}) < \tau_i.$$

□

Remark: Since a simplicial section is in particular a lower-layer corollary 2.18 shows that lemmas 2.11 and 2.12 are indeed correct for an absolutely continuous measure μ.

2.4 Asymptotical Solution

2.4.1 Preparatory lemmas

These two lemmas will be needed in the proof of the main result in the next subsection.

Lemma 2.19 Let $A \subset Q$ be such that $c_i(A) \le s$ for $(i = 1, 2, \ldots, d)$.
Define $A' := \{\mathbf{x} \in A \,|\, \mathbf{x} \le s^{2/3}\mathbf{1}\}$. Then

$$\frac{\mu(A')}{\mu(A)} \ge 1 - d\, s^{1/3} \tag{58}$$

Remark: *We will need this lemma for small s and a simplicial section $A = \Delta$. See also figure 6 on page 42.*
Proof. For $i = 1, \ldots, d$ we have

$$\begin{aligned} s &\ge c_i(A) \\ &= \frac{1}{\mu(A)} \int_A x_i \, d\mu(\mathbf{x}) \\ &\ge \frac{1}{\mu(A)} \int_{A \cap \{x_i \ge s^{2/3}\}} x_i \, d\mu(\mathbf{x}) \\ &\ge \frac{1}{\mu(A)} s^{2/3} \, \mu(A \cap \{x_i \ge s^{2/3}\}). \end{aligned}$$

This implies

$$\frac{\mu(A \cap \{x_i \ge s^{2/3}\})}{\mu(A)} \le s^{\frac{1}{3}}. \tag{59}$$

Now, we get

$$\begin{aligned} 1 - \frac{\mu(A')}{\mu(A)} &= \frac{\mu(A \setminus A')}{\mu(A)} \\ &\le \sum_{i=1}^{d} \frac{\mu(A \cap \{x_i \ge s^{2/3}\})}{\mu(A)} \\ &\le d\, s^{1/3} && \text{(by 59).} \end{aligned}$$

□

Lemma 2.20 (a Chernoff bound) Let $Z_1, Z_2, \ldots, Z_n$ be independent random variables with $E\,[Z_1] \le \frac{1}{n}$ and

$$0 \le Z_i \le a.$$

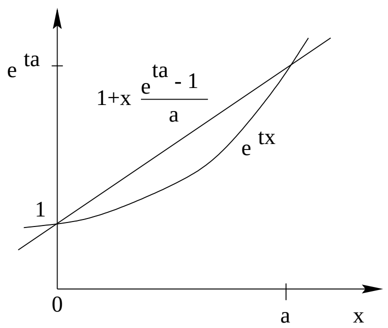
Then we have for any natural number $m < n$

$$P(\sum_{i=1}^{m} Z_i > 1) \leq \exp\,[-\frac{\delta^2}{2a}]$$

with $\delta := 1 - \frac{m}{n}$.

Proof. We have for any $t > 0$

$$\begin{aligned} p := P(\sum_{i=1}^{m} Z_i > 1) &\leq e^{-t} E\,[\exp(t\sum_{i=1}^{m} Z_i)] && \text{(Markov's inequality)} \\ &= e^{-t}\prod_{i=1}^{m} E\,[e^{tZ_i}]. && \text{(independence)} \end{aligned}$$



Observe that, as the function $x \mapsto e^{tx}$ is convex, we have $e^{tx} \leq 1 + x\frac{1}{a}(e^{ta} - 1)$ for $0 \leq x \leq a$.

Plug in Z_i for x to get

$$\begin{aligned} p &\leq e^{-t}\prod_{i=1}^{m}[1 + \frac{\mathrm{E}\,[Z_i]}{a}(e^{ta} - 1)] \\ &= e^{-t}\,[1 + \tfrac{1}{an}(e^{ta} - 1)]^m && \text{(since } \mathrm{E}\,[Z_i] \leq \tfrac{1}{n}) \\ &\leq \exp(-t)\,\exp(\tfrac{m}{an}(e^{ta} - 1)) && \text{(since } 1 + x \leq e^x) \\ &\leq \exp\,\big[\frac{1}{a}((1-\delta)(e^{ta} - 1) - ta)\big] \end{aligned}$$

Since $t > 0$ was arbitrary we can put $t = -\frac{1}{a}\ln(1-\delta) > 0$ and get

$$\begin{aligned} p &\leq \exp\,\big[\frac{1}{a}((1-\delta)(\frac{1}{1-\delta} - 1) + \ln(1-\delta))\big] \\ &\leq \exp\,\big[\frac{1}{a}(\delta + \ln(1-\delta))\big] \end{aligned}$$

Using the power series of ln we get for $0 \leq x < 1$

$$\ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \cdots \leq -x - \frac{x^2}{2}$$

Apply this to $x = \delta$ to get

$$\begin{aligned} p &\leq \exp\big[\frac{1}{a}(\delta - \delta - \frac{\delta^2}{2})\big] \\ &= \exp\big[-\frac{\delta^2}{2a}\big] \end{aligned}$$

□

2.4.2 Result

We will derive asymptotic results about Opt_n for $n \to \infty$ in this section. This is the main result of this chapter.
Again, consider the prophet who is to select the variables. He knows the values of $X_1, \ldots, X_n$ in advance. In this case of fixed n his best policy is clear ; he could 'simply' select the largest subset I of $\{1, 2, \ldots, n\}$ such that $\sum_{i \in I} X_i \leq \mathbf{1}$. The expected number of selected variables by the prophet then is

$$\mathrm{Proph}_n := E\,[\max\{\#I \mid \sum_{i \in I} X_i \leq \mathbf{1}\}].$$

We know $\mathrm{Opt}_n \leq \mathrm{Proph}_n$ but it turns out that for large n his policy is not much better than the optimal online selection policy:

Theorem 2.21 Let μ be an absolutely continuous probability measure on $Q = [\mathbf{0}, \mathbf{1}]$. Let $\Delta = \Delta(n)$ be a simplicial section $\{\mathbf{x} \in Q \mid \langle \mathbf{x}, \boldsymbol{\theta} \rangle \leq 1\}$ such that $g_i(\Delta) \leq 1/n$ and $\theta_i = 0$ when $g_i(\Delta) < 1/n$. Then

(1)
$$\mathrm{Opt}_n \sim \mathrm{Proph}_n \sim n\,\mu(\Delta).$$

(2) Let
$$s := \frac{1}{n\mu(\Delta)} \tag{60}$$
and $\Delta' := \{\mathbf{x} \in \Delta \mid \mathbf{x} \leq s^{2/3}\mathbf{1}\}$. Let Ψ be the following policy (which depends on n). Accept X_j if $X_j \in \Delta'$ and the sum of the variables selected so far plus X_j is still less than or equal to $\mathbf{1}$. But if this sum exceeds $\mathbf{1}$ in any coordinate then reject X_j and all subsequent variables $X_{j+1}, X_{j+2}, \ldots$.
Then Ψ is asymptotically optimal, i.e. $\mathcal{E}(\Psi) \sim \mathrm{Opt}_n$.
Furthermore, for any $\varepsilon > 0$ we have the error bound
$$1 - \frac{\mathcal{E}(\Psi)}{n\,\mu(\Delta)} = O\left(s^{1/3-\varepsilon}\right) \qquad \text{as } n \to \infty. \tag{61}$$

Note: Theorem 2.17 ensures that a Δ like above always exists.

Proof. For the upper bound on Opt_n we will give an upper bound on Proph_n which we can reduce to the one-dimensional case. For the lower bound we will show that Ψ asymptotically achieves the upper bound.

Upper Bound. Recall that, when choosing the variables X_i with $i \in I$, we had to comply with the constraint

$$\sum_{i\in I} X_i \leq \mathbf{1}. \tag{62}$$

Define $\delta := (\theta_1 + \theta_2 + \cdots + \theta_d)^{-1}$ and set $\boldsymbol{\alpha} := \delta\,\boldsymbol{\theta}$. Now consider the following one-dimensional relaxation of (62)

$$\sum_{i\in I} \langle X_i, \boldsymbol{\alpha}\rangle \leq 1. \tag{63}$$

(62) implies (63) because if (62) holds we get

$$\begin{aligned}
\sum_{i\in I} \langle X_i, \boldsymbol{\alpha}\rangle &= \left\langle \sum_{i\in I} X_i, \boldsymbol{\alpha}\right\rangle \\
&= \sum_{j=1}^{d} \alpha_j \underbrace{\sum_{i\in I} X_i^{(j)}}_{\leq 1} \\
&\leq \sum_{j=1}^{d} \alpha_j \\
&= 1.
\end{aligned}$$

Intuitively, (63) means that – instead of staying in the cube Q – the sum of the selected points must stay within a certain simplex given by a hyperplane that goes through $\mathbf{1}$.
Now, let

$$Y_i := \langle X_i, \boldsymbol{\alpha}\rangle\,.$$

When selecting the Y_i's under the relaxed constraint $\sum_{i\in I} Y_i \leq 1$ the prophet will do at least as good as under (62). Let F be the distribution function of the Y_i's. Then F is continuous because μ has a density.
We apply corollary 4.2 to the sequence $(Y_1, Y_2, \ldots)$ to get for $n > 1/\mathrm{E}\,[Y_1]$

$$\mathrm{Proph}_n \leq nF(\varepsilon), \tag{64}$$

for any ε such that

$$\int_0^{\varepsilon} x\,dF(x) = \frac{1}{n}. \tag{65}$$

We will show that $\varepsilon = \delta$ satisfies (65) and that $F(\delta) = \mu(\Delta)$, then (64) implies the upper bound $\text{Proph}_n \lesssim n\mu(\Delta)$.
The latter is clear because

$$F(\delta) = P(Y_1 \leq \delta) = P(\langle X_1, \boldsymbol{\alpha} \rangle \leq \delta) = P(\langle X_1, \boldsymbol{\theta} \rangle \leq 1) = P(X_1 \in \Delta) = \mu(\Delta).$$

Now, show that $\varepsilon = \delta$ satisfies (65).

$$\begin{aligned}
\int_0^\delta x\, dF(x) &= E\,[Y_1 \mathbb{1}_{\{Y_1 \leq \delta\}}] \\
&= E\,[\langle X_1, \boldsymbol{\alpha} \rangle \ \mathbb{1}_{\{\langle X_1, \boldsymbol{\alpha} \rangle \leq \delta\}}] \\
&= E\,[\sum_{j=1}^{d} \alpha_j X_1^{(j)} \mathbb{1}_{\{X_1 \in \Delta\}}] \\
&= \sum_{j=1}^{d} \alpha_j E\,[X_1^{(j)} \mathbb{1}_{\{X_1 \in \Delta\}}] \\
&= \sum_{j=1}^{d} \alpha_j g_j(\Delta) \\
&= \sum_{j=1}^{d} \alpha_j \frac{1}{n} \qquad\qquad \text{because } g_j(\Delta) = \tfrac{1}{n} \text{ if } \alpha_j \neq 0 \\
&= \frac{1}{n}
\end{aligned}$$

We get the upper bound on Opt_n

$$\text{Opt}_n \leq \text{Proph}_n \lesssim n\mu(\Delta). \tag{66}$$

Remark: *As we know already that* $\text{Opt}_n \to \infty$ *by lemma 2.5 we can now conclude that* $n\mu(\Delta) \to \infty$ *as well and therefore*

$$s \to 0 \qquad \text{as } n \to \infty \tag{67}$$

Lower Bound. First note that Ψ as defined in the theorem is an admissible online selection policy. And so the optimal expected number of selected items is at least $\mathcal{E}(\Psi)$:

$$\text{Opt}_n \geq \mathcal{E}(\Psi).$$

In this part we will show the error bound (61). Since $s \to 0$ when $n \to \infty$ this will give us $\mathcal{E}(\Psi) \sim n\mu(\Delta)$. And together with the upper bound we get

$$n\mu(\Delta) \sim \mathcal{E}(\Psi) \leq \text{Opt}_n \leq \text{Proph}_n \lesssim n\mu(\Delta),$$

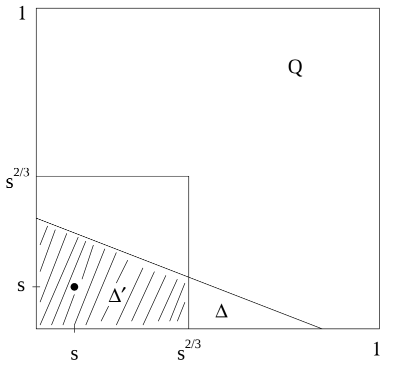


Figure 6: With the policy Ψ a point is selected when it is in the hatched region Δ'. (s is the largest coordinate of the barycenter of Δ.)

which proves the rest of the claim.

The stationary policy which uses Δ instead of Δ' as acceptance region seems to be more natural. Unfortunately, it is not always asymptotically optimal when $d > 1$. But the difference in measure between the two regions is asymptotically negligible:
$c_i(\Delta) = g_i(\Delta)/\mu(\Delta) \leq 1/(n\mu(\Delta)) = s$, so lemma 2.19 gives us

$$\frac{\mu(\Delta')}{\mu(\Delta)} \geq 1 - d\, s^{1/3}, \tag{68}$$

which converges to 1 as $n \to \infty$. We will need (68) later.

The upper bound on $\mathcal{E}(\Psi)$ is easy. Trivially, we have

$$\begin{aligned} \mathcal{E}(\Psi) &\leq E\,[\sum_{i=1}^{n} \mathbb{1}_{\{X_i \in \Delta'\}}] \\ &= n\,\mu(\Delta') \\ &\leq n\,\mu(\Delta). \end{aligned}$$

So

$$1 - \frac{\mathcal{E}(\Psi)}{n\,\mu(\Delta)} \geq 0.$$

Now, we turn to the lower bound on $\mathcal{E}(\Psi)$.
Introduce the stopping time (see definition 4.7)

$$\rho := \inf\{k \mid \sum_{i=1}^{k} X_i \mathbb{1}_{\{X_i \in \Delta'\}} \not\leq \mathbf{1}\}$$

and set $\rho = \infty$ if no such k exists. Then the number of selected variables by our strategy Ψ is

$$\sum_{i=1}^{(\rho-1)\wedge n} \mathbb{1}_{\{X_i \in \Delta'\}},$$

where $(\rho-1)\wedge n$ denotes the minimum of n and $\rho-1$. When using $\rho \wedge n$, which is also a stopping time, instead of $(\rho-1)\wedge n$ as upper bound of the sum the error is at most 1 and will be asymptotically negligible. Both $\rho \wedge n$ and $\mathbb{1}_{\{X_i \in \Delta'\}}$ have finite expectation, whence by Wald's equation (see theorem 4.8),

$$\begin{aligned} \mathcal{E}(\Psi) &\geq E\,[\sum_{i=1}^{\rho\wedge n} \mathbb{1}_{\{X_i \in \Delta'\}}] - 1 \\ &= \mu(\Delta') \cdot E\,[\rho \wedge n] - 1. \end{aligned} \tag{69}$$

For $\mu(\Delta')$ we already have a bound. Now, we want to bound $E[\rho \wedge n]$ from below. We will use that

$$E\,[\rho \wedge n] \geq m\,P(\rho > m) \qquad \text{for any } m < n \tag{70}$$

and will have to choose m suitably. Note that

$$\begin{aligned} P(\rho \leq m) &= P(\sum_{i=1}^{m} X_i \mathbb{1}_{\{X_i \in \Delta'\}} \not\leq \mathbf{1}) \\ &\leq \sum_{j=1}^{d} P(\sum_{i=1}^{m} X_i^{(j)} \mathbb{1}_{\{X_i \in \Delta'\}} > 1) \end{aligned}$$

Now, apply lemma 2.20 to $Z_i = X_i^{(j)} \mathbb{1}_{\{X_i \in \Delta'\}}$, $a = s^{2/3}$ and $m = \lfloor n(1 - s^{1/3-\varepsilon}) \rfloor$. This is possible because $\mathrm{E}\,[Z_i] = g_j(\Delta') \leq g_j(\Delta) \leq \frac{1}{n}$ and $Z_i \leq a$ by definition of Δ'.
We get with $\delta := 1 - \frac{m}{n} \geq s^{1/3-\varepsilon}$

$$\begin{aligned} P(\rho \leq m) &\leq d\,\exp\,[-\frac{\delta^2}{2a}] \\ &\leq d\,\exp\,[-\frac{s^{2/3-2\varepsilon}}{2s^{2/3}}] \\ &= d\,\exp\,[-\frac{1}{2}s^{-2\varepsilon}] \\ &= \mathrm{O}(s) \qquad\qquad \text{as } s \to 0. \end{aligned} \tag{71}$$

Now we are ready to prove the rest. Take inequality (69) to start with.

$$\begin{aligned} \frac{\mathcal{E}(\Psi)}{n\,\mu(\Delta)} &\geq \frac{\mu(\Delta')\mathrm{E}\,[\rho \wedge n] - 1}{n\,\mu(\Delta)} \\ &\geq \frac{\mu(\Delta')}{\mu(\Delta)}\,\frac{m}{n}\,P(\rho > m)\ - s \qquad\qquad \text{(by inequality (70))} \end{aligned}$$

All three factors on the right-hand side are less than 1 (and converge to 1) and we already have bounds for them. First write

$$1 - \frac{\mathcal{E}(\Psi)}{n\,\mu(\Delta)} \le 1 - \frac{\mu(\Delta')}{\mu(\Delta)}\,\frac{m}{n}\,P(\rho > m)\; + s$$

and now use the inequality $1 - abc \le (1-a) + (1-b) + (1-c)$, which holds for all $0 \le a, b, c \le 1$ and apply (68), the definition of m and (71) to get

$$\begin{aligned} 1 - \frac{\mathcal{E}(\Psi)}{n\,\mu(\Delta)} &\le ds^{\frac{1}{3}} \;+\; \mathrm{O}(s^{1/3-\varepsilon}) \;+\; \mathrm{O}(s) + s \\ &= \mathrm{O}(s^{1/3-\varepsilon}). \end{aligned}$$

□

2.5 Solution for certain distributions

2.5.1 A product distribution

Consider a measure μ on Q with distribution function

$$F(\mathbf{x}) = a x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_d^{\alpha_d}$$

in a neighborhood of $\mathbf{0}$, where $a, \alpha_1, \alpha_2, \ldots, \alpha_d$ are positive constants. This is the direct generalisation of the distributions studied in the paper of Coffman and al. [3] to more than one dimension. This will also give the result for the Lebesgue-measure on Q as a special case.

Theorem 2.22

$$\mathrm{Opt}_n \sim \gamma \cdot (an)^{1/(1+\alpha)} \tag{72}$$

where

$$\gamma := (1+\alpha)\left[\frac{\alpha_1 \cdots \alpha_d\, \Gamma(\alpha_1) \cdots \Gamma(\alpha_d)}{\alpha_1^{\alpha_1} \cdots \alpha_d^{\alpha_d}\, \Gamma(2+\alpha)}\right]^{1/(1+\alpha)}$$

and $\alpha := \alpha_1 + \alpha_2 + \cdots + \alpha_d$.

Proof. The proof is given in section 3.2. This is a special case of theorem 3.2 where the problem with random N is treated.

Having this solution we can derive an asymptotic solution to an **integro-differential equation**.
The special form of the distribution function F enables us to use a different approach to obtain Opt_n in this case. Again, consider the Poissonized 'version' of this problem. Let $u(t)$ be as in subsection 2.2.2. Similar to the recursion formula for Opt_n in section 2.1 we will get an integro-differential equation for $u(t)$. The solution to the equation will be unique so we get information about it via $u(t)$.

Lemma 2.23 Let
$$F(\mathbf{x}) = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_d^{\alpha_d}$$
be the distribution function of the measure μ. Let f be a density of μ with respect to the Lebesgue measure. Then u is differentiable and satisfies the integro-differential equation

$$u'(t) = \int_Q (1 + u(F(\mathbf{1} - \mathbf{y})t) - u(t))^+ \, dF(\mathbf{y}). \tag{73}$$

Proof. Suppose at some point of the selection process we are left with the space $\mathbf{c} \leq \mathbf{1}$, that is to say the sum of the items selected so far is $\mathbf{1} - \mathbf{c}$. And suppose there is still t time left. We want to express the maximal expected number of selected items in this 'rest'-problem, i.e $u(\mu, \mathbf{c}, t)$, by u.
Since we cannot select items with sizes which are not in $[\mathbf{0}, \mathbf{c}]$ we might as well consider the 'thinned' counting process of those events j of the original Poisson process such that $X_j \in [\mathbf{0}, \mathbf{c}]$. It is well known that this new process then again is a homogenuous Poisson process with rate $P(X_j \in [\mathbf{0}, \mathbf{c}]) = F(\mathbf{c})$. It is equivalent to again consider a Poisson process with rate 1 but shorten the time of the selection process to $F(\mathbf{c})\, t$ because in both cases the events come from a homogenuous Poisson process and their total number has the same distribution (Poisson with mean $F(\mathbf{c})\, t$).
Let μ' be the probability measure on $[\mathbf{0}, \mathbf{c}]$ defined by $\mu'(A) = \mu(A)/F(\mathbf{c})$ for subsets $A \subset [\mathbf{0}, \mathbf{c}]$. Then μ' is the distribution of X_j given that $X_j \in [\mathbf{0}, \mathbf{c}]$. So $u(\mu, \mathbf{c}, t) = u(\mu', \mathbf{c}, F(\mathbf{c})t)$.
We will now show that the packing problem on $[\mathbf{0}, \mathbf{c}]$ with measure μ' is equivalent to the packing problem on $[\mathbf{0}, \mathbf{1}]$ with measure μ. Here, equivalent means $u(\mu', \mathbf{c}, s) = u(\mu, \mathbf{1}, s)$ $(s \geq 0)$.
Consider the linear, one-to-one mapping

$$\begin{aligned} g : [\mathbf{0}, \mathbf{1}] &\rightarrow [\mathbf{0}, \mathbf{c}] \\ (x_1, x_2, \ldots, x_d) &\mapsto (c_1 x_1, c_2 x_2, \ldots, c_d x_d). \end{aligned}$$

which respects the partial order in $\mathbb{R}^d$. g is measure-preserving in the sense

$$\mu(g^{-1}(A)) = \mu'(A) \tag{74}$$

for all $A \subset [\mathbf{0}, \mathbf{c}]$. To prove this, it suffices to show (74) for $A = [\mathbf{0}, \mathbf{y}]$ and every $0 \leq \mathbf{y} \leq \mathbf{c}$ because the the set of all such intervals is $\cap$-stable and generates the σ-algebra $\mathcal{B}([\mathbf{0}, \mathbf{c}])$. So $g(\mu) = \mu'$ then follows by the uniqueness theorem in measure theory. But

$$\mu(g^{-1}(A)) = \mu(\,[\mathbf{0}, g^{-1}(\mathbf{y})]\,) = F(g^{-1}(\mathbf{y})) = \frac{F(\mathbf{y})}{c_1^{\alpha_1} \cdots c_d^{\alpha_d}} = \frac{F(\mathbf{y})}{F(\mathbf{c})} = \mu'(A).$$

By the use of g we can now transform an admissible policy Ψ for the packing problem on $[\mathbf{0}, \mathbf{1}]$ with measure μ to an admissible policy Ψ' for the packing problem on $[\mathbf{0}, \mathbf{c}]$ with measure μ' and the other way around. If Ψ is given define

$$\Psi'(\,(x_1, t_1),\, (x_2, t_2), \ldots) = \Psi(\,(g^{-1}(x_1), t_1),\, (g^{-1}(x_2), t_2), \ldots)$$

and if Ψ' is given define

$$\Psi(\,(x_1,t_1),\,(x_2,t_2),\ldots) = \Psi'(\,(g(x_1),t_1),\,(g(x_2),t_2),\ldots).$$

As g is measure-preserving we have $\mathcal{E}_\mu(\Psi) = \mathcal{E}_{\mu'}(\Psi')$.

Therefore our problem is quivalent to packing on $[\mathbf{0},\mathbf{1}]$ with the items coming from a Poisson process with rate 1 in time $F(c)\,t$ and we get $u(\mu,\mathbf{c},t) = u(\mu,\mathbf{1},F(\mathbf{c})t) = u(F(\mathbf{c})t)$ as maximal expected number of selected variables.

Now, to give a formula for $u(t+\varepsilon)$ for some small ε, we condition on the number of events occurring in the time $[0,\varepsilon]$: $L(\varepsilon)$. $L(\varepsilon)$ is Poisson distributed with mean ε. Therefore $P(L(\varepsilon)=0) = 1-\varepsilon-o(\varepsilon)$, $\quad P(L(\varepsilon)=1) = \varepsilon+o(\varepsilon)$ and $P(L(\varepsilon)\geq 2) = o(\varepsilon)$ (see definition (4.5)). If no event occurs up to time ε then the maximal expected number of selected items will be $u(t)$. If one event occurs up to time ε and $X_1 = \mathbf{y}$ then we have the choice of rejecting or selecting it. If we reject it we get a maximal expected number of selected items of $u(t)$. If we select it, we get this one item plus the maximal expected number of selected items we can select in time t and space $\mathbf{1}-\mathbf{y}$. By the dynamical programming principle, it is optimal to choose the larger value. Thus by the above discussion we have a maximal expected number of selected items of

$$\int_Q f(\mathbf{y})\ \max\{1+u(F(\mathbf{1}-\mathbf{y})t), u(t)\}\,d\mathbf{y},$$

conditioned on $\{1$ event occurs up to time $\varepsilon\}$. Since the probability of two or more events occurring up to time ε becomes so small it suffices to note that this conditional maximal expected number of selected items is finite.
All in all we get

$$\begin{aligned} u(t+\varepsilon) \;=\;& (1-\varepsilon-o(\varepsilon))\,u(t) \\ & +(\varepsilon+o(\varepsilon))\int_Q f(\mathbf{y})\ \max\{1+u(F(\mathbf{1}-\mathbf{y})t), u(t)\}\,d\mathbf{y} \\ & +o(\varepsilon). \end{aligned}$$

Collect the terms with $o(\varepsilon)$, rearrange and divide by ε to obtain

$$\frac{u(t+\varepsilon)-u(t)}{\varepsilon} = -u(t) + \int_Q f(\mathbf{y})\ \max\{1+u(F(\mathbf{1}-\mathbf{y})t), u(t)\}\,d\mathbf{y} + \frac{o(\varepsilon)}{\varepsilon}$$

Since $o(\varepsilon)/\varepsilon \to 0$ when $\varepsilon \to 0$ we see that u is differentiable and satisfies

$$\begin{aligned} u'(t) &= -u(t) + \int_Q f(\mathbf{y})\ \max\{1+u(F(\mathbf{1}-\mathbf{y})t), u(t)\}\,d\mathbf{y} \\ &= \int_Q f(\mathbf{y})\,(\max\{1+u(F(\mathbf{1}-\mathbf{y})t), u(t)\} - u(t))\,d\mathbf{y} \\ &= \int_Q f(\mathbf{y})\,(1+u(F(\mathbf{1}-\mathbf{y})t) - u(t))^+\,d\mathbf{y} \end{aligned}$$

□

Plugging f and F into this equation, yields the following integro-differential equation with initial condition.

$$\left.\begin{array}{l} u'(t) = \alpha_1 \cdots \alpha_d \int_Q y_1^{\alpha_1 - 1} \cdots y_d^{\alpha_d - 1} \left[1 + u((1-y_1)^{\alpha_1} \cdots (1-y_d)^{\alpha_d}\, t) - u(t)\right]^+ d\mathbf{y} \\ u(0) = 0 \end{array}\right\} \tag{75}$$

Corollary 2.24 The solution u to the integro-differential equation (75) is unique and satisfies

$$u(t) \sim \gamma \cdot t^{1/(1+\alpha)} \qquad (t \to \infty).$$

Proof. We know that u as defined in (11) is a solution to (75). Since $u(n) \sim \mathrm{Opt}_n$ by theorem 2.8, we get $u(t) \sim \gamma t^{1/(1+\alpha)}$ by theorem 2.22. The only thing that is left to prove is the uniqueness of the solution to (75).
Suppose $u \neq v$ were two different solutions. Then let $b := \inf\{t \geq 0 \,|\, u(t) \neq v(t)\}$.
Then $u(b) = v(b)$ because u and v are continuous and $u(0) = v(0)$. Let $\varepsilon > 0$. By integrating the integro-differential equation and using $u(b) = v(b)$ we get

$$u(b+\varepsilon) - v(b+\varepsilon) = \int_b^{b+\varepsilon} \int_Q f(\mathbf{y})[(1+u(F(1-\mathbf{y})t) - u(t))^+ - (1+v(F(1-\mathbf{y})t) - v(t))^+]\, d\mathbf{y}\, dt$$

Using $|x^+ - y^+| \leq |x - y|$ we get

$$\begin{aligned} |u(b+\varepsilon) - v(b+\varepsilon)| &\leq \int_b^{b+\varepsilon} \int_Q f(\mathbf{y})[|u(F(1-\mathbf{y})t) - v(F(1-\mathbf{y})t) + v(t) - u(t)|]\, d\mathbf{y}\, dt \\ &\leq \varepsilon\, 2\, \underbrace{\max\{|u(t) - v(t)| \,|\, 0 \leq t \leq b+\varepsilon\}}_{=:m}, \end{aligned}$$

where the last step follows by the triangle inequality, since $F(1-\mathbf{y}) \leq 1$ and since $\mu(Q) = 1$. Now, let $\varepsilon < \frac{1}{2}$ be such that $m = |u(b+\varepsilon) - v(b+\varepsilon)|$, that is the maximum is attained at the right border of $[0, b+\varepsilon]$. It is always possible to choose such an ε by making the interval smaller. As $m > 0$ we get the contradiction $m < m$.

□

2.5.2 'Almost' Lebesgue measure

Let the density f be continuous in 0 and let

$$a := f(0) > 0$$

Then

$$\mathrm{Opt}_n \sim \gamma \cdot (an)^{\frac{1}{d+1}}$$

with

$$\gamma = \frac{d+1}{((d+1)!)^{1/(d+1)}}.$$

Remark: *As the solution to the selection problem asymptotically only depends on the behavior of $f(\mathbf{x})$ when $\mathbf{x} \to \mathbf{0}$ this distribution can be regarded as 'almost' Lebesgue measure up to the scaling factor a.*

Proof. This is a special case of the example in the previous section. Set $\alpha_1 = \alpha_2 = \cdots = \alpha_d = 1$. Then the density of the distribution in example 1 is constant a in a neighborhood of $\mathbf{0}$. And $f(\mathbf{x}) \sim a$ when $\mathbf{x} \to \mathbf{0}$. So, by theorem 2.9 we get that Opt_n is asymptotically like the expression (72) from theorem 2.22.

$$\mathrm{Opt}_n \sim \gamma \cdot (an)^{\frac{1}{d+1}}.$$

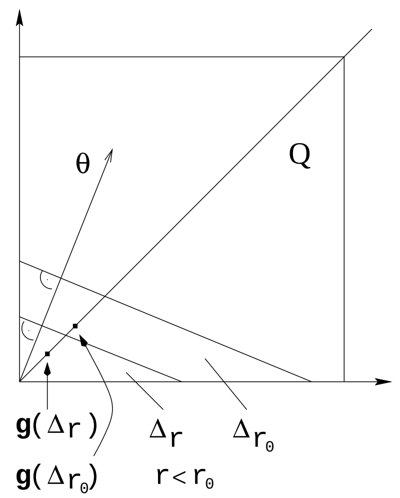
Plugging $\alpha_i = 1$ into the expression for γ and using $\Gamma(1) = 1$ verifies the above result.

□

3 Sequential Selection out of a Random Number of Random Vectors under a Sum Constraint

3.1 Main result for the random number of items

In this chapter we again have a random number of items at disposal to be packed into $Q = [0,1]^d$. We cannot take a fixed acceptance region anymore to get an asymptotically optimal policy. For instance if N is geometrically distributed, intuitively we would choose a large acceptance region in the beginning. After all, with fixed probability the item presented to us could be the last and space that is left over doesn't help us. And later, when the remaining space is small, we would choose – or even have to choose – a smaller acceptance region.
Again, we only solve the problem asymptotically when ν becomes 'large' in a sense. This problem is more genereral than the one from the previous chapter but we get the increase in generality only by using a stronger hypothesis on the measure μ on Q:



There must be a $\boldsymbol{\theta} \in R_+^d$, $\boldsymbol{\theta} > \mathbf{0}$ such that

$$g_1(\Delta_r) = g_2(\Delta_r) = \cdots = g_d(\Delta_r)$$

with

$$\Delta_r := \{\mathbf{x} \in Q \mid \langle \mathbf{x}, \boldsymbol{\theta} \rangle \leq r\}$$

for all r sufficiently small ($0 < r \leq r_0$). (76)

(76) means that the simplicial sections Δ_r, which are similar to one another in the geometric sense, must have $\mathbf{g}(\Delta_r)$ on the diagonal of the positive orthant at least when they are small (see the figure to the left).

Remark: *This seems like a strong hypothesis. It is true that theorem 2.17 suggests that most absolutely continuous measures admit finding simplicial sections with* $\mathbf{g}$ *on the diagonal but in general they need not to be similar.*
On the other hand, in the following section we will show that the fairly general μ *with the distribution function*

$$P(X_1 \leq \mathbf{x}) = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_d^{\alpha_d},$$

where $\alpha_1, \alpha_2, \ldots, \alpha_d$ *are positive constants, does satisfiy this hypothesis.*
There are also distributions μ *satisfying (76), where the coordinates* $X_1^{(1)}, X_1^{(2)}, \ldots, X_1^{(d)}$ *are dependent. Every exchangeable* μ *also satisfies (76) with* $\theta = \mathbf{1}$. *Exchangeable means that the distribution of* $(X_1^{(\sigma(1))}, X_1^{(\sigma(2))}, \ldots, X_1^{(\sigma(d))})$ *is the same for all permutations* σ *of* $\{1, 2, \ldots, d\}$.

Let $\boldsymbol{\theta}$ be as above, without loss of generality we can norm $\boldsymbol{\theta}$ such that $\|\boldsymbol{\theta}\|_1 = 1$ and define

$$S_i := \langle \boldsymbol{\theta}, X_i \rangle .$$

Let F be the distribution function of the S_i's. And let

$$G(x) = \mathrm{E}\left[S_1 \mathbb{1}_{\{S_1 \le x\}}\right] = \int_0^x s\, dF(s). \tag{77}$$

Further suppose that

$$\begin{array}{l} F \text{ is continuous and} \\ xF(x) \le c\, G(x) \text{ for some constant } c > 0 \text{ and for all } 0 < x < x_0, \end{array} \tag{78}$$

where x_0 can be arbitrarily small. Note that then G must be continuous also.
Define

$$\pi_j := P(N \ge j) \qquad\qquad \text{for } j = 1, 2, \ldots .$$

And let

$$\mathrm{Opt}_\nu := \mathrm{Opt}(\mu, \mathbf{1}, \nu).$$

Theorem 3.1 Suppose the measure μ on Q satisfies (76) and (78).

(a) If ν – the distribution of N – varies in such a way that the equation

$$\sum_{j=1}^{\infty} G(\varepsilon \pi_j) = 1 \tag{79}$$

can be solved for ε and $\varepsilon \to 0$,
then

$$\mathrm{Opt}_\nu \sim \sum_{j=1}^{\infty} \pi_j F(\varepsilon \pi_j) \to \infty. \tag{80}$$

The strategy $\hat{\Psi}$ with

$$\hat{\Psi}_j(X_1, \ldots, X_j) = 1 \quad :\Longleftrightarrow \quad S_j \le \varepsilon \pi_j \quad \text{and} \quad \sum_{i=1}^{j} X_i \mathbb{1}_{\{S_i \le \varepsilon \pi_j\}} \le \mathbf{1}$$

is asymptotically optimal, i.e. $\mathcal{E}(\hat{\Psi}) \sim \mathrm{Opt}_\nu$.
Furthermore, we have the error bound

$$\frac{\mathcal{E}(\hat{\Psi})}{\mathrm{Opt}_\nu} = 1 - \mathsf{O}(\varepsilon^{\frac{1}{3}}). \tag{81}$$

(b) If $\sum_{j=1}^{\infty} G(\varepsilon \pi_j) > 1$, for all $\varepsilon > 0$ then $\mathrm{Opt}_\nu = \infty$.

(c) If $\sum_{j=1}^{\infty} G(\varepsilon\pi_j) < 1$, for all $\varepsilon > 0$ then $\operatorname{ess\,sup} N < 1/\mathrm{E}[S_1]$.

Remark: *We don't make a statement about* Opt_ν *if* $\varepsilon \not\to 0$ *or in case (c). In these cases* ν *is 'not large enough' to allow an asymptotical analysis by our method.*

Proof. Let $\theta_{\min} := \min_i \theta_i > 0$.

(a): In part 1 we show the upper bound and in part 2 we show that above $\hat{\Psi}$ asymptotically achieves this bound. But first note that the expression in the claim (80) indeed tends to infinity when $\varepsilon \to 0$.

$$\begin{aligned} \sum_{j=1}^{\infty} \pi_j F(\varepsilon\pi_j) &= \frac{1}{\varepsilon} \sum_{j=1}^{\infty} \varepsilon\pi_j \, F(\varepsilon\pi_j) \\ &\geq \frac{1}{\varepsilon} \sum_{j=1}^{\infty} G(\varepsilon\pi_j) && \text{since } xF(x) \geq G(x) \\ &= \frac{1}{\varepsilon} \to \infty \end{aligned}$$

Part 1 (Upper Bound). We will show that the upper bound not only holds asymptotically but

$$\mathrm{Opt}_\nu \leq \sum_{j=1}^{\infty} \pi_j F(\varepsilon\pi_j). \tag{82}$$

First extend the class of admissible policies. Let $\mathcal{P}'$ be the class of all selection policies Ψ (they need not to be online!) which satisfy the one dimensional sum constraint

$$\sum_{j=1}^{\infty} \Psi_j S_j \leq 1. \tag{83}$$

The one dimensional sum constraint (83) is weaker than the sum constraint (1), since (1) implies

$$\sum_{j=1}^{\infty} \Psi_j S_j = \sum_{j=1}^{\infty} \Psi_j \langle X_j, \boldsymbol{\theta} \rangle = \Big\langle \underbrace{\sum_{j=1}^{\infty} \Psi_j X_j}_{\leq \mathbf{1} \text{ by } (1)}, \boldsymbol{\theta} \Big\rangle \leq \theta_1 + \cdots + \theta_d = 1.$$

And therefore

$$\mathrm{Opt}_\nu \leq \sup_{\Psi \in \mathcal{P}'} \mathcal{E}(\Psi) = \mathrm{Proph}(\mathrm{dist}(S_1), 1, \nu).$$

For the latter we have an upper bound given by theorem 4.1. Apply theorem 4.1 to the sequence $(S_1, S_2, \ldots)$, the distribution of S_1 and with N, F and ε like above. This is possible since

$$\mathrm{E}\sum_{j=1}^{\infty} S_1 \mathbb{1}_{\{S_1 \leq \varepsilon\pi_j\}} = \sum_{j=1}^{\infty} G(\varepsilon\pi_j) = 1,$$

so hypothesis (103) is satisfied.
We get

$$\mathrm{Proph}(\mathrm{dist}(S_1), 1, \nu) \leq \sum_{j=1}^{\infty} \pi_j F(\varepsilon\pi_j),$$

which proves (82).

Part 2 (Lower Bound). Let Ψ' be defined by $\Psi'_j = \mathbb{1}_{\{S_j \leq \varepsilon\pi_j\}}$.
Let $\hat{\Psi}$ be the policy from the theorem, i.e

$\hat{\Psi}_j = 1 :\Longleftrightarrow \Psi'_j = 1$ *and* the sum of the variables selected so far plus X_j lies still in Q.

The first time this sum leaves Q the selection process stops and all further random variables are rejected. Since the decision whether to select X_j depends only on $X_1, \ldots, X_j$ and since (1) is fulfilled the selection policy $\hat{\Psi}$ lies in $\mathcal{P}$.
Now let

$$\rho := \min\{k \mid \sum_{j=1}^{k} \Psi'_j X_j \not\leq \mathbf{1}\}$$

and let $\rho := \infty$ if no such k exists. Then

$$\mathcal{E}(\hat{\Psi}) = \mathrm{E}\sum_{j=1}^{(\rho-1)\wedge N} \Psi'_j. \tag{84}$$

Since

$$\mathrm{E}\sum_{j=1}^{N} \Psi'_j = \sum_{j=1}^{\infty} \mathrm{E}\mathbb{1}_{\{N \geq j\}} = \sum_{j=1}^{\infty} \pi_j F(\varepsilon\pi_j)$$

we have to show that

$$\lim_{\varepsilon\to 0} \frac{\mathrm{E}\sum_{j=1}^{(\rho-1)\wedge N} \Psi'_j}{\mathrm{E}\sum_{j=1}^{N} \Psi'_j} = 1 \tag{85}$$

and bound the error. In this form we see that we must show that the stopping time ρ isn't 'too often too small'. In order to be able to use an independence-argument later in the proof we would like to use $\rho \wedge N$ instead of $(\rho - 1) \wedge N$ as upper bound of the sum. The denominator tends to infinity as $\varepsilon \to 0$ as proven in the very beginning. For this reason and since E $\sum_{j=1}^{(\rho-1)\wedge N} \Psi'_j$ and E $\sum_{j=1}^{\rho\wedge N} \Psi'_j$ differ by at most 1, we don't change the limit (and neither the order of magnitude of the error bound) when using the latter nominator:

$$\begin{aligned} E\sum_{j=1}^{\rho\wedge N} \Psi'_j &= \sum_{j=1}^{\infty} E\,\Psi'_j \mathbb{1}_{\{N\geq j\}} \mathbb{1}_{\{\rho\geq j\}} \\ &= \sum_{j=1}^{\infty} F(\varepsilon\pi_j)\ \mathrm{P}(N\geq j)\ \mathrm{P}(\rho\geq j) \qquad (86) \\ &= \sum_{j=1}^{\infty} \pi_j F(\varepsilon\pi_j) - \sum_{j=1}^{\infty} \pi_j\, F(\varepsilon\pi_j)\, P(\rho < j), \qquad (87) \end{aligned}$$

where the line next to the last ,(86), follows from the independence of N, Ψ'_j and the event $\{\rho \geq j\} = \{X_1 + \cdots + X_{j-1} \leq \mathbf{1}\}$. The expression in (87) makes sense ($\infty - \infty$ is impossible) as will turn out later. Since the first term in (87) is what we would like to approximate, we need to bound the second term from above. We have for any $J \in \mathbb{N}$ (which we will choose properly in the sequel)

$$\begin{aligned} \sum_{j=1}^{\infty} \pi_j F(\varepsilon\pi_j) P(\rho<j) &= \sum_{j=1}^{J} \pi_j F(\varepsilon\pi_j) P(\rho<j) + \sum_{j=J+1}^{\infty} \pi_j F(\varepsilon\pi_j) P(\rho<j) \\ &\leq P(\rho\leq J) \sum_{j=1}^{J} \pi_j F(\varepsilon\pi_j) + \sum_{j=J+1}^{\infty} \pi_j F(\varepsilon\pi_j) \\ &\leq P(\rho\leq J)\, \frac{1}{\varepsilon} \sum_{j=1}^{\infty} \varepsilon\pi_j\, F(\varepsilon\pi_j) + \frac{1}{\varepsilon} \sum_{j=J+1}^{\infty} \varepsilon\pi_j\, F(\varepsilon\pi_j) \\ &\leq P(\rho\leq J) \frac{1}{\varepsilon} c + \frac{1}{\varepsilon} c\delta, \qquad (88) \end{aligned}$$

where last line holds because $xF(x) \leq cG(x)$, $\sum_{j=1}^{\infty} G(\varepsilon\pi_j) = 1$ and with

$$\delta = \delta(J) := \sum_{j=J+1}^{\infty} G(\varepsilon\pi_j).$$

Now, bound $P(\rho \leq J)$ from above.

$$\begin{aligned} P(\rho\leq J) &= P\big(\sum_{j=1}^{J} X_j \mathbb{1}_{\{S_j\leq\varepsilon\pi_j\}} \not\leq \mathbf{1}\big) \\ &\leq \sum_{i=1}^{d} \mathrm{P}\big(\sum_{j=1}^{J} X_j^{(i)} \mathbb{1}_{\{S_j\leq\varepsilon\pi_j\}} > 1\big) \qquad (89) \end{aligned}$$

We will bound the probability on the right-hand side with Chebyshev's inequality. First we have to compute $\mathrm{E}X_j^{(i)} 1\!\!1_{\{S_j \le \varepsilon\pi_j\}}$. Observe that $\{S_j \le r\} = \{\langle X_j, \boldsymbol{\theta}\rangle \le r\} = \{X_j \in \Delta_r\}$. So that $\mathrm{E}X_j^{(i)} 1\!\!1_{\{S_j \le \varepsilon\pi_j\}} = g_i(\Delta_{\varepsilon\pi_j})$. We claim that

$$g_i(\Delta_r) = G(r) \qquad\qquad \text{for } 0 < r < r_0. \tag{90}$$

Let $0 < r < r_0$. Then $g_i(\Delta_r)$ is the same for all $i = 1, 2, \ldots, d$ by the hypothesis (76) and we get

$$\begin{aligned} G(r) &= \mathrm{E}\, S_1 1\!\!1_{\{S_1 \le r\}} = \mathrm{E}\ S_1 1\!\!1_{\{X_1 \in \Delta_r\}} = \left\langle \mathrm{E}\, X_1 1\!\!1_{\{X_1 \in \Delta_r\}}, \boldsymbol{\theta} \right\rangle \\ &= \langle \mathbf{g}(\Delta_r), \boldsymbol{\theta}\rangle = (\theta_1 + \theta_2 + \cdots + \theta_d)\, g_1(\Delta_r) = g_1(\Delta_r), \end{aligned}$$

which proves (90).

But since $\varepsilon\pi_j \le \varepsilon < r_0$ for ε small enough we conclude that then $\mathrm{E}\, X_j^{(i)} 1\!\!1_{\{S_j \le \varepsilon\pi_j\}} = G(\varepsilon\pi_j)$.

Using (79) and the definition of δ we get

$$\mathrm{E} \sum_{j=1}^{J} X_j^{(i)} 1\!\!1_{\{S_j \le \varepsilon\pi_j\}} = \sum_{j=1}^{J} G(\varepsilon\pi_j) = 1 - \delta. \tag{91}$$

Now bound the variance using the independence of the X_j's and the formula $\mathrm{VAR}[Z] \le \mathrm{E}\,[Z^2]$ for a random variable Z.

$$\mathrm{VAR} \sum_{j=1}^{J} X_j^{(i)} 1\!\!1_{\{S_j \le \varepsilon\pi_j\}} \le \sum_{j=1}^{\infty} \mathrm{E}\, (X_j^{(i)})^2 1\!\!1_{\{S_j \le \varepsilon\pi_j\}} \tag{92}$$

On the event $\{S_j \le \varepsilon\pi_j\}$ we have $\theta_i X_j^{(i)} \le \langle \boldsymbol{\theta}, X_j\rangle = S_j \le \varepsilon\pi_j \le \varepsilon$ and thus $X_j^{(i)} \le \frac{\varepsilon}{\theta_i}$. We bound one of the factors of the square on the right-hand side of (92) by $\frac{\varepsilon}{\theta_i}$ and get

$$\begin{aligned} \mathrm{VAR} \sum_{j=1}^{J} X_j^{(i)} 1\!\!1_{\{S_j \le \varepsilon\pi_j\}} &\le \frac{\varepsilon}{\theta_i} \sum_{j=1}^{\infty} \mathrm{E}\, X_j^{(i)} 1\!\!1_{\{S_j \le \varepsilon\pi_j\}} \\ &= \frac{\varepsilon}{\theta_i} \sum_{j=1}^{\infty} G(\varepsilon\pi_j) \\ &= \frac{\varepsilon}{\theta_i} \end{aligned} \tag{93}$$

Finally, we can apply Chebyshev's inequality to (89) using (91) and (93) to get

$$P(\rho \le J) \le \sum_{i=1}^{d} \frac{1}{\delta^2} \frac{\varepsilon}{\theta_i} \le \frac{\varepsilon d}{\delta^2 \theta_{\min}}. \tag{94}$$

Substituting this result in (88) and using $\sum\limits_{j=1}^{\infty} \pi_j \, F(\varepsilon\pi_j) \geq \frac{1}{\varepsilon}$ yields

$$\frac{\sum\limits_{j=1}^{\infty} \pi_j \, F(\varepsilon\pi_j) \, P(\rho < j)}{\sum\limits_{j=1}^{\infty} \pi_j \, F(\varepsilon\pi_j)} \leq \frac{dc}{\delta^2 \frac{1}{\varepsilon}\theta_{\min}} + c\delta \leq \text{const} \cdot (\frac{\varepsilon}{\delta^2} + \delta) \tag{95}$$

for all J.
Now, we have to choose $J(\varepsilon)$, such that this term tends to zero, when $\varepsilon \to 0$. Recall that

$$\delta = \sum_{j=J+1}^{\infty} G(\varepsilon\pi_j).$$

Since the sum converges by the hypothesis and every term in the sum is less than or equal to ε (since $G(x) \leq x$,) we can choose J such that $\delta \sim \varepsilon^{\frac{1}{3}}$. Then

$$\frac{\varepsilon}{\delta^2} + \delta = \mathrm{O}(\varepsilon^{\frac{1}{3}}).$$

Putting together (95) and (87) we obtain

$$\frac{E \sum\limits_{j=1}^{\rho\wedge N} \Psi'_j}{\sum_{j=1}^{\infty} \pi_j F(\varepsilon\pi_j)} = 1 - \mathrm{O}(\varepsilon^{\frac{1}{3}}),$$

which indeed remains valid if we again write $\rho - 1$ instead of ρ, since $\varepsilon = \mathrm{O}(\varepsilon^{\frac{1}{3}})$.
(84) and (82) now give the result

$$\frac{\mathcal{E}(\hat{\Psi})}{\mathrm{Opt}_\nu} = 1 - \mathrm{O}(\varepsilon^{\frac{1}{3}}).$$

(b): We will show that for arbitrary large $M \in \mathbb{N}$ there is a strategy Ψ such that $\mathcal{E}(\Psi) \geq M$. Then $\mathrm{Opt}_\nu = \infty$ follows.
First, we will show that (b) even implies

$$\sum_{j=1}^{\infty} G(\varepsilon\pi_j) = \infty \qquad \text{for all } \varepsilon > 0. \tag{96}$$

Suppose for $\varepsilon = \varepsilon_1 > 0$ this series was finite, say $\leq L$. Then we can find a J such that $\sum_{j=J+1}^{\infty} G(\varepsilon_1\pi_j) \leq L/4$ and an ε_2, $0 < \varepsilon_2 < \varepsilon_1$ such that $\sum_{j=1}^{J} G(\varepsilon_2\pi_j) \leq L/4$. The latter is possible since the finite sum is continuous in ε_2 and vanishes for $\varepsilon_2 = 0$. Because G is increasing, we get $\sum_{j=1}^{\infty} G(\varepsilon_2\pi_j) \leq L/4 + L/4 = L/2$. We could continue halving the

value of the series until it drops below 1. Contradiction.
Fix $M \in \mathbb{N}$ and let $\varepsilon > 0$ be such that

$$\varepsilon \leq \frac{\theta_{\min}}{2M} \quad \text{and } \varepsilon < \frac{1}{6}.$$

Consider the (inadmissible) policy Ψ' with $\Psi'_j := \mathbb{1}_{\{S_j \leq \varepsilon\pi_j\}}$.

This policy might violate the constraint $\sum_{j=1}^{\infty} X_j \Psi'_j \leq \mathbf{1}$. But – as the $\varepsilon\pi_j$ are small – whenever this restriction is violated we must have selected at least $2M$ items before: Suppose we have $\sum_{j=1}^{k} X_j \Psi'_j \not\leq \mathbf{1}$ for some k and the constraint is violated in the i-th coordinate. Then

$$\begin{aligned} 1 &< \sum_{j=1}^{k} X_j^{(i)} \mathbb{1}_{\{S_j \leq \varepsilon\pi_j\}} \\ &\leq \sum_{j=1}^{k} \frac{\varepsilon\pi_j}{\theta_i} \mathbb{1}_{\{S_j \leq \varepsilon\pi_j\}} && \text{, as on } \{S_j \leq \varepsilon\pi_j\} \text{ we have } \theta_i X_j^{(i)} \leq S_j \leq \varepsilon\pi_j \\ &\leq \sum_{j=1}^{k} \frac{1}{2M} \mathbb{1}_{\{S_j \leq \varepsilon\pi_j\}} && \text{, as } \tfrac{\varepsilon\pi_j}{\theta_i} \leq \tfrac{\varepsilon}{\theta_i} \leq 1/(2M). \end{aligned}$$

We conclude that

$$\sum_{j=1}^{k} \mathbb{1}_{\{S_j \leq \varepsilon\pi_j\}} > 2M.$$

So the number of selected items with Ψ' before the first violation of the sum restriction is at least $2M$.
We will show that this violation of the constraint happens at least with probability $\frac{1}{2}$.

$$P(\text{constraint not violated}) \leq P\Big(\sum_{j=1}^{\infty} S_j \mathbb{1}_{\{S_j \leq \varepsilon\pi_j\}} \leq 1\Big) \leq P(S \leq 1)$$

with $S := \sum_{j=1}^{J} S_j \mathbb{1}_{\{S_j \leq \varepsilon\pi_j\}}$ and for any J. By (96) we can choose J such that

$$\mathrm{E}\,[S] = \sum_{j=1}^{J} G(\varepsilon\pi_j)$$

is larger than 2 and smaller than 3. We have

$$\mathrm{VAR}\,[S] \leq \sum_{j=1}^{J} \mathrm{E}\, S_j^2 \mathbb{1}_{\{S_j \leq \varepsilon\pi_j\}} \leq \varepsilon\pi_j \sum_{j=1}^{J} \mathrm{E} S_j \mathbb{1}_{\{S_j \leq \varepsilon\pi_j\}} \leq \varepsilon\,\mathrm{E}\,[S] \leq 3\varepsilon \leq \frac{1}{2},$$

which implies by the Chebyshev inequality that $P(S \leq 1) \leq \mathrm{VAR}\,[S]/(2-1)^2 < 1/2$.

Therefore the policy Ψ which selects all items that Ψ'_j selects as long as the sum constraint is satisfied: $\Psi_j = 1 :\Longleftrightarrow \Psi'_j = 1$ and $\sum_{l=1}^{j} X_l \Psi'_l \leq \mathbf{1}$
has

$$\mathcal{E}(\Psi) \geq 2M \cdot \frac{1}{2} = M.$$

□

(c): Let k be so that $\pi_k > 0$. Then we can define $\varepsilon = \frac{1}{\pi_k}$ and get

$$1 > \sum_{j=1}^{\infty} G(\varepsilon\pi_j) \geq \sum_{j=1}^{k} G(1) = k\mathrm{E}\,[S_1].$$

If $\operatorname{ess\,sup} N$ was ∞ then $\mathrm{E}\,[S_1] = 0$, which we have excluded because the distribution function F of S_1 was continuous. So, letting $k = \operatorname{ess\,sup} N$ we get $\operatorname{ess\,sup} N < 1/\mathrm{E}\,[S_1]$.

□

3.2 Example

Theorem 3.2 Let μ have the distribution function

$$P(X_1 \leq \mathbf{x}) = a x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_d^{\alpha_d}$$

for $\mathbf{x}$ in a neighborhood U of $\mathbf{0}$, where $a, \alpha_1, \alpha_2, \ldots, \alpha_d$ are positive constants.
Write $\alpha := \alpha_1 + \alpha_2 + \cdots + \alpha_d$ and let

$$\gamma := (1+\alpha)\left[\frac{\alpha_1 \cdots \alpha_d\, \Gamma(\alpha_1)\cdots\Gamma(\alpha_d)}{\alpha_1^{\alpha_1}\cdots\alpha_d^{\alpha_d}\,\Gamma(2+\alpha)}\right]^{1/(1+\alpha)}. \tag{97}$$

If

$$n_\pi := \sum_{j=1}^{\infty} \pi_j^{1+\alpha} \to \infty$$

as the distribution of N varies , then

$$\mathrm{Opt}_\nu \sim \gamma \cdot (a n_\pi)^{\frac{1}{1+\alpha}}.$$

If $n_\pi = \infty$ then $\mathrm{Opt}_\nu = \infty$.

Remark: *n_π is a measure for the 'size of N'.*

- *In the case of fixed $N \equiv n$ we have $n_\pi = n$. We have stated this special case in theorem 2.22.*

- *In the case where N is geometric distributed with parameter $p = 1 - q$, i.e. $P(N = j) = p\,q^{j-1}$ for $j = 1, 2, \ldots$, we have $\pi_j = q^{j-1}$ and*

$$n_\pi = \sum_{j=1}^{\infty} (q^{j-1})^{1+\alpha} = \sum_{j=0}^{\infty} (q^{1+\alpha})^j = \frac{1}{1 - q^{1+\alpha}}.$$

We see that $n_\pi \to \infty$ is equivalent to $p \to 0$. And in that case we have $1 - q^{1+\alpha} \sim (1+\alpha)\,p$, which one can see using de l'Hospital's rule. So we have

$$\mathrm{Opt}_\nu \sim \gamma \cdot \left[\frac{a}{(1+\alpha)\,p}\right]^{1/(1+\alpha)}.$$

Proof of theorem 3.2.
Let $\boldsymbol{\theta}$ be defined by $\theta_i := \alpha_i/\alpha$. Then $\boldsymbol{\theta} > \mathbf{0}$ and $\|\boldsymbol{\theta}\|_1 = 1$. We will show that with this choice the hypotheses (76) and (78) of theorem 3.1 are satisfied.
Let $\Delta_r := \{\mathbf{x} \in Q \mid \langle \mathbf{x}, \boldsymbol{\theta} \rangle \le r\}$. Observe that there is a $r_0 > 0$ such that $\Delta_r = \{\mathbf{x} \ge \mathbf{0} \mid \langle \mathbf{x}, \boldsymbol{\theta} \rangle \le r\}$ and $\Delta_r \subset U$ for $0 < r < r_0$. In the following let $0 < r < r_0$.

We calculate

$$\begin{aligned} g_i(\Delta_r) &= \int_{\Delta_r} x_i \, d\mu(\mathbf{x}) \\ &= a \int_{\Delta_r} \alpha_1 \alpha_2 \cdots \alpha_d \quad x_1^{\alpha_1 - 1} x_2^{\alpha_2 - 1} \cdots x_i^{\alpha_i} \cdots x_d^{\alpha_d - 1} \, d\mathbf{x} \end{aligned}$$

For clearness we put the dots here as well as in the rest of the example although the i-th factor is an 'exception'. The dots have to be interpreted as if the i-th factor was not there. Now, we substitute $y_j = \frac{\theta_j}{r} x_j$ ($j = 1, \ldots, d$). Then $\Delta_r = \{\mathbf{y} \ge \mathbf{0} \mid y_1 + \cdots + y_d \le 1\}$. As $\boldsymbol{\theta} > \mathbf{0}$, we get

$$g_i(\Delta_r) = a\, r^{1+\alpha} \, \frac{\alpha_1 \cdots \alpha_d}{\theta_1^{\alpha_1} \cdots \theta_d^{\alpha_d} \theta_i} \int\limits_{\substack{y_1 + \cdots + y_d \le 1 \\ y_j \ge 0}} y_1^{\alpha_1 - 1} \cdots y_i^{\alpha_i} \cdots y_d^{\alpha_d - 1} \, d\mathbf{y}.$$

Now use Dirichlet's formula (e.g. see [5] or [18]), the functional equation $\Gamma(x+1) = x\Gamma(x)$ of the Gamma function and the definition of $\boldsymbol{\theta}$ to calculate the integral.

$$\begin{aligned} g_i(\Delta_r) &= a\, r^{1+\alpha} \, \frac{\alpha_1 \cdots \alpha_d}{\theta_1^{\alpha_1} \cdots \theta_d^{\alpha_d} \theta_i} \; \frac{\Gamma(\alpha_1) \cdots \Gamma(\alpha_i + 1) \cdots \Gamma(\alpha_d)}{\Gamma(1 + \alpha_1 + \cdots + (\alpha_i + 1) + \cdots + \alpha_d)} \\ &= a\, r^{1+\alpha} \, \frac{\alpha_i}{\theta_i} \, \frac{\alpha_1 \cdots \alpha_d}{\theta_1^{\alpha_1} \cdots \theta_d^{\alpha_d}} \, \frac{\Gamma(\alpha_1) \cdots \Gamma(\alpha_d)}{\Gamma(2 + \alpha)} \\ &= a \left(\frac{r\alpha}{1+\alpha} \gamma \right)^{1+\alpha}. \end{aligned} \tag{98}$$

This shows that $g_i(\Delta_r)$ does not depend on i and so (76) is proven.
This also gives us $G(r)$ as defined in (77), because by (90) we have $G(r) = g_i(\Delta_r)$.
So, using (98) we get

$$G(r) = k\, r^{1+\alpha} \qquad \text{with } k := a\, (\frac{\alpha}{1+\alpha}\gamma)^{1+\alpha}. \tag{99}$$

We also need to calculate $F(r)$. In principal we could use a very similar way as when we computed $g_i(\Delta_r)$ with Dirichlet's formula. But as we already have $G(r)$, we can use the definition of G – formula (77) – to calculate F through G:

$$G(r) = \int_0^r x\, dF(x) = [xF(x)]_0^r - \int_0^r F(x)\, dx = rF(r) - \int_0^r F(x)\, dx \tag{100}$$

Here, we made use of partial integration. G is differentiable and F is continuous by definition because μ is continuous. So the integral to the right is differentiable with respect to r. Therefore, rearranging (100) for $F(r)$, we see that $F(r)$ must also be differentiable in $r > 0$.
Differentiate (100) to get

$$G'(r) = rF'(r) + F(r) - F(r) = rF'(r).$$

So

$$\begin{aligned} F(r) &= F(0) + \int_0^r F'(x)\, dx \\ &= \int_0^r \frac{G'(x)}{x}\, dx \\ &= \int_0^r \frac{(1+\alpha)kx^\alpha}{x}\, dx \\ &= \frac{1+\alpha}{\alpha}\, k\, r^\alpha \\ &= \frac{1+\alpha}{\alpha}\, \frac{G(r)}{r}. \end{aligned} \tag{101}$$

This also shows that hypothesis (78) is also satisfied with $c = \frac{1+\alpha}{\alpha}$.
Now consider the equation

$$\sum_{j=1}^{\infty} G(\varepsilon \pi_j) = 1.$$

It is equivalent to

$$1 = k\, \varepsilon^{1+\alpha} \sum_{j=1}^{\infty} \pi_j^{1+\alpha} = k\, n_\pi \varepsilon^{1+\alpha}.$$

So if $n_\pi = \infty$, $\sum_{j=1}^{\infty} G(\varepsilon\pi_j) > 1$ for all $\varepsilon > 0$ and theorem 3.1 tells us that $\mathrm{Opt}_\nu = \infty$.
Otherwise

$$\varepsilon = (kn_\pi)^{-1/(1+\alpha)}, \tag{102}$$

which tends to 0 as n_π tends to ∞. We can apply theorem 3.1 and get

$$
\begin{aligned}
\mathrm{Opt}_\nu &\sim \sum_{j=1}^{\infty} \pi_j F(\varepsilon\pi_j) = \sum_{j=1}^{\infty} \pi_j \tfrac{1+\alpha}{\alpha} \tfrac{G(\varepsilon\pi_j)}{\varepsilon\pi_j} && \text{(using (101))} \\
&= \tfrac{1+\alpha}{\alpha} \tfrac{1}{\varepsilon} \sum_{j=1}^{\infty} G(\varepsilon\pi_j) \\
&= \frac{1+\alpha}{\alpha}\frac{1}{\varepsilon} \\
&= \tfrac{1+\alpha}{\alpha} (kn_\pi)^{1/(1+\alpha)} && \text{(by (102))} \\
&= \gamma \cdot (an_\pi)^{1/(1+\alpha)} && \text{(by definition of } k\text{)}
\end{aligned}
$$

□

4 Basic Theorems Used in the Main Part

Section 4.1 contains a one-dimensional result about the maximal expected number of selected variables by the prophet, which we have used both in chapter 2 and 3.
The other section of this chapter constitutes of 'well known' facts.

4.1 Sequential Selection under a Constraint in one Dimension

Theorem 4.1 Let $X_1, X_2, \ldots \geq 0$ be a sequence of independent identically distributed random variables with distribution μ and distribution function F. Let $N \sim \nu$ be a random variable with values in $\mathbb{N}$ that is independent of $(X_1, X_2, \ldots)$. Let $\pi_j := P(N \geq j)$ and $\varepsilon > 0$ be such that

$$\mathrm{E} \sum_{j=1}^{\infty} X_1 \mathbb{1}_{\{X_1 \leq \varepsilon \pi_j\}} = 1. \tag{103}$$

Then the expected number of items the prophet can pack is

$$\mathrm{Proph}(\mu, 1, \nu) \leq \sum_{j=1}^{\infty} \pi_j F(\varepsilon \pi_j). \tag{104}$$

Proof. The proof is much like in [8], where the idea of using the expected sum constraint (105) comes from.
Let $\mathcal{Q}$ be the class of all selection policies Ψ that satisfy

$$\mathrm{E} \sum_{j=1}^{\infty} X_j \Psi_j \leq 1. \tag{105}$$

Clearly, (105) is weaker than the sum constraint (1). So

$$\mathrm{Proph}(\mu, 1, \nu) \leq \sup_{\Psi \in \mathcal{Q}} \mathcal{E}(\Psi). \tag{106}$$

We claim that the following section policy Ψ' (which is even online!) defined by

$$\Psi'_j := \begin{cases} 1 & , \text{ if } X_j \leq \varepsilon \pi_j \\ 0 & , \text{ if } X_j > \varepsilon \pi_j \end{cases} \tag{107}$$

is optimal in $\mathcal{Q}$. That is $\Psi' \in \mathcal{Q}$ and $\mathcal{E}(\Psi') = \sup_{\Psi \in \mathcal{Q}} \mathcal{E}(\Psi)$.
To proof this, define

$$g(\Psi) = \mathrm{E} \sum_{j=1}^{\infty} \Psi_j X_j \ - 1$$

for any selection policy Ψ. Then the expected sum constraint (105) is equivalent to $g(\Psi) \le 0$.
To prove $\Psi' \in \mathcal{Q}$ we only have to calculate $g(\Psi)$:

$$\begin{aligned} g(\Psi') &= \mathrm{E}\sum_{j=1}^{\infty} \Psi'_j X_j \ - 1 \\ &= \mathrm{E}\sum_{j=1}^{\infty} X_j \mathbb{1}_{\{X_j \le \varepsilon\pi_j\}} \ - 1 \\ &= 0 \qquad \text{(by hypothesis)} \end{aligned}$$

Now, we show that $\mathcal{E}(\Psi') \ge \mathcal{E}(\Psi)$ for all $\Psi \in \mathcal{Q}$. For this purpose define the Lagrangian

$$\begin{aligned} L(\Psi) &:= \mathcal{E}(\Psi) - \frac{1}{\varepsilon} g(\Psi) \qquad (108) \\ &= \mathrm{E}\sum_{j=1}^{N} \Psi_j - \frac{1}{\varepsilon} \cdot \big(\mathrm{E}\sum_{j=1}^{\infty} \Psi_j X_j \ - 1\big) \\ &= \frac{1}{\varepsilon}\Big(1 + \sum_{j=1}^{\infty} \mathrm{E}\left[\varepsilon \Psi_j \mathbb{1}_{\{N \ge j\}} - \Psi_j X_j\right]\Big) \\ &= \frac{1}{\varepsilon}\Big(1 + \sum_{j=1}^{\infty} \mathrm{E}\left[\Psi_j(\varepsilon\pi_j - X_j)\right]\Big) \qquad \text{since } N \text{ and } \Psi \text{ are independent.} \end{aligned}$$

In this form we see that $L(\Psi)$ is maximized over all selection policies when we choose

$$\begin{aligned} \Psi_j = 1 &\quad \text{, if } \ \varepsilon\pi_j - X_j \ge 0 \text{ and} \\ \Psi_j = 0 &\quad \text{, if } \ \varepsilon\pi_j - X_j < 0. \end{aligned}$$

And this is equivalent to the definition (107) of Ψ'. So

$$L(\Psi') \ge L(\Psi) \qquad \text{for all selection policies } \Psi. \qquad (109)$$

We conclude that for arbitrary $\Psi \in \mathcal{Q}$ we have

$$\begin{aligned} \mathcal{E}(\Psi) &\le L(\Psi) && \text{by (108), since } g(\Psi) \le 0 \\ &\le L(\Psi') && \text{by (109)} \\ &= \mathcal{E}(\Psi') && \text{by (108), since } g(\Psi') = 0 \end{aligned}$$

So

$$\sup_{\Psi \in \mathcal{Q}} \mathcal{E}(\Psi) = \mathcal{E}(\Psi') = \mathrm{E}\sum_{j=1}^{N} \mathbb{1}_{\{X_j \le \varepsilon\pi_j\}} = \sum_{j=1}^{\infty} \pi_j F(\varepsilon\pi_j).$$

Together with (106) this proves the claim (104).

□

Corollary 4.2 If in theorem 4.1 $N \equiv n$ is not random and we have $n > 1/\mathrm{E}[X_1]$ then there is an $\varepsilon > 0$ such that $G(\varepsilon) = \frac{1}{n}$ with

$$G(x) := \int_0^x y\, dF(y).$$

And we have

$$\mathrm{Proph}_n(\mu, 1) \leq nF(\varepsilon)$$

for any such ε.

Proof. Note that in this case $\pi_j = 1$ if $j \leq n$ and $\pi_j = 0$ otherwise, so we have

$$\mathrm{E}\sum_{j=1}^{\infty} X_1 1\!\!1_{\{X_1 \leq \varepsilon\pi_j\}} = n\mathrm{E}X_1 1\!\!1_{\{X_1 \leq \varepsilon\}} = nG(\varepsilon),$$

where we used that $P(X_1 = 0) = 0$, since F is continuous. Since G must be continuous also, $G(0) = 0$ and $G(\varepsilon) \uparrow \mathrm{E}X_1 > \frac{1}{n}$ as $\varepsilon \to \infty$ there is an ε such that $nG(\varepsilon) = 1$. The preceeding theorem tells us that

$$\mathrm{Proph}_n(\mu, 1) \leq \sum_{j=1}^{\infty} \pi_j F(\varepsilon\pi_j) = nF(\varepsilon).$$

□

4.2 Standard theorems

Kuhn-Tucker theorem

Definition 4.3 Let H be a real vector space, Y a normed space and f be a (possibly nonlinear) transformation from H to Y defined on a domain $D \subset H$.
Let $x \in D$ and let h be arbitrary in H. If the limit

$$\delta f(x; h) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}[f(x + \varepsilon h) - f(x)] \tag{110}$$

exists, it is called the *Gateaux-variation* of f with *increment* h. If the limit (110) exists for all $h \in H$, the transformation f is said to be *Gateaux differentiable* at x.
A point at which $\delta f(x; h) = 0$ for all h is called a *stationary point.*

Theorem 4.4 (Generalized Kuhn-Tucker Theorem) Let H be a real vector space, f be a Gateaux differentiable real-valued functional on H and g a Gateaux differentiable mapping from H into $\mathbb{R}^d$. Assume that the Gateaux-variations are linear in their increments. Suppose $\hat{x}$ minimizes f subject to $g(x) \leq 0$ and that $\hat{x}$ is a regular point of the inequality $g(x) \leq 0$, i.e. there is an $h \in X$ such that $g(\hat{x}) + \delta g(\hat{x}, h) < 0$. Then there is a $\boldsymbol{\theta} \in \mathbb{R}^d$, $\boldsymbol{\theta} \geq \mathbf{0}$, such that the Lagrangian

$$f(x) + \langle g(x), \boldsymbol{\theta} \rangle$$

is stationary at $\hat{x}$; furthermore, $\langle g(\hat{x}), \boldsymbol{\theta} \rangle = 0$.

This was taken from [11].

Poisson process

The following definition is taken from [15].

Definition 4.5 (Poisson process) The counting process $\{L(t),\ t \geq 0\}$ is said to be a Poisson process with rate a, $a > 0$, if:

1. $L(0) = 0$.
2. The process has stationary and independent increments.
3. $P(L(h) = 1) = a\,h + o(h)$.
4. $P(N(h) \geq 2) = o(h)$.

It is well known that $L(t)$ is Poisson distributed with mean at. Also

Theorem 4.6 (coloring, thinning) Suppose in a Poisson process $\{L(t),\ t \geq 0\}$ having rate a every event is colored blue with probability p and red with probability $1-p$ independently of the rest. Then the counting processes $\{B(t),\ t \geq 0\}$ and $\{R(t),\ t \geq 0\}$ of the blue and red events are independent Poisson processes having respective rates ap and $a(1-p)$.

Wald's Equation

This definition and theorem are taken from [15].

Definition 4.7 An integer-valued random variable ρ is said to be a *stopping time* for the sequence $X_1, X_2, \ldots$ if the event $\{\rho = k\}$ is independent of $X_{k+1}, X_{k+2}, \ldots$ for all $k = 1, 2, \ldots$.

Theorem 4.8 (Wald's Equation)
If $X_1, X_2, \ldots$ are independent and identically distributed random variables having finite expectations, and if ρ is a stopping time for $X_1, X_2, \ldots$ such that $\mathrm{E}\,[\rho] < \infty$, then

$$\mathrm{E} \sum_{j=1}^{\rho} X_j = \mathrm{E}\,[\rho]\,\mathrm{E}\,[X_1]$$

Topology and Analysis

Theorem 4.9 (Lemma of Urysohn in strong form) Let A and B be closed disjoint G_δ subsets of the normal topological space X. Then there exists a function $f : X \to [0,1]$ such that $f^{-1}(0) = A$ and $f^{-1}(1) = B$.

Note: If X is metric than X is normal and every closed subset is G_δ.
See [13].

Theorem 4.10 (Arzelà-Ascoli) Let (S, d) be a compact metric space and let $M \subset C(S)$. Where $C(S)$ – the set of all continuous functions from S to $\mathbb{R}$ – is endowed with the supremum norm. Suppose

(a) M is bounded and

(b) M is equicontinuous, i.e.

$$\forall \varepsilon > 0 \, \exists \delta > 0 \, \forall x \in M : \quad d(s,t) < \delta \Longrightarrow |x(s) - x(t)| < \varepsilon.$$

Then M is relatively compact, i.e. every sequence in M has a subsequence which converges in $C(S)$.

See [16].

References

[1] Bauer, Heinz (1990): *Maß- und Integrationstheorie*, Walter de Gruyter

[2] Bruss, F. Thomas, Robertson, James B. (1991): *'Wald's Lemma' for Sums of Order Statistics of i.i.d Random Variables*, Adv. Appl. Prob. **23**, 612-623

[3] Coffman, E.G., Flatto, L. and Weber, R.R. (1987): *Optimal selection of stochastic intervals under a sum constraint.*, Adv. Appl. Prob. **19**, 454-473.

[4] Coffman, E.G. jr., Lueker, G.S. (1991): *Probabilistic Analysis of Packing and Partitioning Algorithms*, John Wiley & Sons

[5] Dirichlet, G. Lejeune- (1904) , *Vorlesungen über die Lehre von den einfachen und mehrfachen bestimmten Integralen*, Friedrich Vieweg und Sohn, Braunschweig

[6] Fichtenholz, G.M. (1970): *Differential- und Integralrechnung III*, 3. Auflage, VEB Deutscher Verlag der Wissenschaften, Berlin

[7] Gnedin, Alexander V., Baryshnikov, Yuliy M. (1999): *Sequential Selection of an Increasing Sequence from a Multidimensional Random Sample*, to appear in *Annals of Applied Probability*

[8] Gnedin, Alexander V. (1999): *Sequential Selection under a Sum Constraint when the Number of Observations is Random*, to appear in *Proc. 4th Hungarian Conf. on Limit Theorems in Probability and Statistics*

[9] Gnedin, Alexander V. (1991): *Sequential Selection of an Increasing Subsequence from a Random Sample with Geometric Sample-Size*, preprint

[10] Gnedin, Alexander V. (1991): *Sequential Selection of an Increasing Subsequence from a Sample of Random Size*

[11] Luenberger, David G. *Optimization by Vector Space Methods* John Wiley & Sons

[12] Mallows C.L., Nair V.N., Shepp L.A. and Vardi Y. (1985): *Optimal Sequential Selection of Secretaries*, Mathematics of Operations Research, **10**, No. 4.

[13] Munkres, James R. (1975): *Topology - a first course*, Prentice-Hall, New Yersey

[14] Rhee, Wansoo, Talagrand, Michel (1991) *A Note on the Selection of Random Variables under a Sum Constraint*, J. Appl. Prob. **28**, 919-923

[15] Ross, Sheldon M. (1996) *Stochastic Processes*, John Wiley & Sons

[16] Werner, Dirk (1997) *Funktionanalysis* Springer, Berlin, Heidelberg, New York

[17] Werner, Jochen (1984): *Optimization Theory and Applications*, Vieweg Advanced Lectures in Mathematics

[18] Whittaker, E.T., Watson, G.N. (1996) *A Course of Modern Analysis*, Cambridge University Press